

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 11:13:25 ; Search time 22.2651 Seconds
(without alignments)
83.400 Million cell updates/sec

Title: US-10-783-095-11
Perfect score: 28
Sequence: 1 LAKEMQALCAYQAEPTCATPQEGNIK 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgm2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgm2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgm2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgm2_6/ptodata/1/iaa/5B_COMB.pep.*
5: /cgm2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgm2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	100.0	294	US-08-468-576B-18	Sequence 18, Appl
2	28	100.0	294	US-08-468-579B-18	Sequence 18, Appl
3	28	100.0	294	US-08-468-577B-18	Sequence 18, Appl
4	28	100.0	548	US-08-468-576B-19	Sequence 19, Appl
5	28	100.0	548	US-08-468-579B-19	Sequence 19, Appl
6	28	100.0	548	US-08-468-577B-19	Sequence 19, Appl
7	28	100.0	979	US-08-514-213A-2	Sequence 2, Appl
8	28	100.0	979	US-09-015-399-5	Sequence 5, Appl
9	24	85.7	24	US-09-552-802B-22	Sequence 22, Appl
10	19	67.9	19	US-09-552-802B-20	Sequence 20, Appl
11	18	64.3	17	US-09-552-802B-21	Sequence 21, Appl
12	17	60.7	18	US-09-552-802B-19	Sequence 19, Appl
13	16	57.1	16	US-09-552-802B-18	Sequence 18, Appl
14	13	46.4	13	US-09-552-802B-17	Sequence 17, Appl
15	10	35.7	298	US-08-811-481-4	Sequence 4, Appl
16	10	35.7	298	US-09-876-527-4	Sequence 4, Appl
17	10	35.7	376	US-08-811-481-7	Sequence 7, Appl
18	10	35.7	376	US-09-876-527-7	Sequence 7, Appl
19	10	35.7	641	US-08-811-481-2	Sequence 2, Appl
20	10	35.7	641	US-09-876-527-2	Sequence 2, Appl
21	10	35.7	723	US-08-548-159-5	Sequence 5, Appl
22	10	35.7	818	US-08-811-481-22	Sequence 22, Appl
23	10	35.7	818	US-09-876-527-22	Sequence 22, Appl
24	10	35.7	969	US-08-548-159-1	Sequence 1, Appl
25	10	35.7	986	US-08-548-159-3	Sequence 3, Appl
26	10	35.7	1001	US-08-884-569A-2	Sequence 2, Appl
27	10	35.7	1012	US-08-811-481-16	Sequence 16, Appl

28	10	35.7	1012	4	US-09-876-527-16
29	8	28.6	8	4	US-09-552-802B-49
30	6	21.4	98	4	US-09-489-039A-9193
31	6	21.4	98	4	US-09-489-039A-9959
32	6	21.4	98	4	US-09-489-039A-10543
33	6	21.4	98	4	US-09-489-039A-11154
34	6	21.4	98	4	US-09-489-039A-11319
35	6	21.4	98	4	US-09-489-039A-11469
36	6	21.4	98	4	US-09-489-039A-11572
37	6	21.4	98	4	US-09-489-039A-11873
38	6	21.4	98	4	US-09-489-039A-12042
39	6	21.4	98	4	US-09-489-039A-12617
40	6	21.4	204	6	5187077-22
41	6	21.4	217	4	US-09-270-767-43882
42	6	21.4	402	4	US-09-270-767-61551
43	6	21.4	432	4	US-09-252-991A-24690
44	6	21.4	478	5	PCT-US95-08493-15
45	6	21.4	530	5	PCT-US95-08493-2

ALIGNMENTS

RESULT 1
US-08-468-576B-18
; Sequence 18, Application US/08468576B
; Patent No. 5955345
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,576B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.7-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids

Db US-08-468-576B-18
 TYPE: amino acid
 TOPOLOGY: linear
 Query Match 100.0%; Score 28; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 6.3e-22;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 67 LAKEWQALCAVQAEPNTCATAQGEGNK 94

Qy 1 LAKEWQALCAVQAEPNTCATAQGEGNK 28
 |||||
 Db 67 LAKEWQALCAVQAEPNTCATAQGEGNK 94
 |||||

RESULT 3
 US-08-468-577B-18
 ; Sequence 18, Application US/08468577B
 ; Patent No. 6001804
 ; GENERAL INFORMATION:
 ; APPLICANT: Rabin, Daniel
 ; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
 ; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
 ; STREET: 660 White Plains Road
 ; CITY: Tarrytown
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10591-5144
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: System 7.5
 ; SOFTWARE: WordPerfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/468,577B
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/239,276
 ; FILING DATE: 05-MAY-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/872,646
 ; FILING DATE: 08-JUN-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/312,543
 ; FILING DATE: 17-FEB-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kurt G. Briscoe
 ; REGISTRATION NUMBER: 33,141
 ; REFERENCE/DOCKET NUMBER: MDI 251.8-KGB
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (914) 332-1700
 ; TELEFAX: (914) 332-1844
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 294 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-08-468-577B-18

RESULT 4
 US-08-468-576B-19
 ; Sequence 19, Application US/08468576B
 ; Patent No. 5955345
 ; GENERAL INFORMATION:
 ; APPLICANT: Rabin, Daniel
 ; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS

Query Match 100.0%; Score 28; DB 3; Length 294;
 Best Local Similarity 100.0%; Pred. No. 6.3e-22;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LAKEWQALCAVQAEPNTCATAQGEGNK 28
 |||||
 Db 67 LAKEWQALCAVQAEPNTCATAQGEGNK 94
 |||||

RESULT 4
 US-08-468-576B-19
 ; Sequence 19, Application US/08468576B
 ; Patent No. 5955345
 ; GENERAL INFORMATION:
 ; APPLICANT: Rabin, Daniel
 ; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS

US-08-468-579B-18
 ; Sequence 18, Application US/08468579B
 ; Patent No. 5981700
 ; GENERAL INFORMATION:
 ; APPLICANT: Rabin, Daniel
 ; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
 ; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
 ; STREET: 660 White Plains Road
 ; CITY: Tarrytown
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10591-5144
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: System 7.5
 ; SOFTWARE: WordPerfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/468,579B
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/239,276
 ; FILING DATE: 05-MAY-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/872,646
 ; FILING DATE: 08-JUN-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/715,181
 ; FILING DATE: 14-JUN-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: MDI 251.5-KGB
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (914) 332-1700
 ; TELEFAX: (914) 332-1844
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 294 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-08-468-579B-18

Query Match 100.0%; Score 28; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 6.3e-22;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LAKEWQALCAVQAEPNTCATAQGEGNK 28
 |||||
 Db 67 LAKEWQALCAVQAEPNTCATAQGEGNK 94
 |||||

RESULT 4
 US-08-468-576B-19
 ; Sequence 19, Application US/08468576B
 ; Patent No. 5955345
 ; GENERAL INFORMATION:
 ; APPLICANT: Rabin, Daniel
 ; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 11:13:25 ; Search time 8.74699 Seconds
(without alignments)
83.400 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 11
Sequence: 1 QPLALEGSLQK 11

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_AA:*
1: /sgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /sgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /sgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /sgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /sgn2_6/ptodata/1/iaa/FACTUS_COMB.pep:*
6: /sgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	11	100.0	23	US-08-105-904B-10	Sequence 10, Appl
2	11	100.0	23	US-08-114-877A-10	Sequence 10, Appl
3	11	100.0	23	US-08-537-928-10	Sequence 10, Appl
4	11	100.0	23	US-09-633-269-10	Sequence 10, Appl
5	11	100.0	27	US-08-508-664-17	Sequence 17, Appl
6	11	100.0	29	US-09-676-787-2	Sequence 2, Appl
7	11	100.0	32	5514646-5	Patent No. 5514646
8	11	100.0	33	US-09-639-667-24	Sequence 24, Appl
9	11	100.0	33	5514646-4	Patent No. 5514646
10	11	100.0	35	US-08-160-376A-3	Sequence 3, Appl
11	11	100.0	35	US-08-389-487-6	Sequence 6, Appl
12	11	100.0	35	US-09-134-836-3	Sequence 3, Appl
13	11	100.0	35	US-09-386-303A-3	Sequence 3, Appl
14	11	100.0	35	US-09-815-229-4	Sequence 4, Appl
15	11	100.0	35	US-09-947-563-3	Sequence 3, Appl
16	11	100.0	35	5457066-1	Patent No. 5457066
17	11	100.0	54	US-09-815-229-13	Sequence 13, Appl
18	11	100.0	86	US-07-715-183C-1	Sequence 1, Appl
19	11	100.0	86	US-09-477-924-2	Sequence 2, Appl
20	11	100.0	86	US-09-723-981-2	Sequence 2, Appl
21	11	100.0	86	US-09-723-896-2	Sequence 2, Appl
22	11	100.0	86	US-09-878-380-1	Sequence 1, Appl
23	11	100.0	90	US-08-030-73A-43	Sequence 43, Appl
24	11	100.0	91	US-09-676-787-7	Sequence 7, Appl
25	11	100.0	96	US-09-134-836-4	Sequence 4, Appl
26	11	100.0	96	US-09-134-836-5	Sequence 5, Appl
27	11	100.0	96	US-09-386-303A-4	Sequence 4, Appl

Sequence 5, Appl
Sequence 4, Appl
Sequence 5, Appl
Sequence 4, Appl
Sequence 7, Appl
Sequence 6, Appl
Sequence 7, Appl
Sequence 8, Appl
Sequence 11, Appl
Sequence 7, Appl
Sequence 8, Appl
Sequence 9, Appl
Sequence 11, Appl
Sequence 2, Appl
Sequence 4, Appl
Sequence 2, Appl
Sequence 4, Appl
Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-105-904B-10
; Sequence 10, Application US/08105904B
; Patent No. 6001364
; GENERAL INFORMATION:
; APPLICANT: Rose, Keith
; APPLICANT: Offord, Robin
; TITLE OF INVENTION: HETERO-POLYOXIME COMPOUNDS AND THEIR
; PREPARATION BY PARALLEL ASSEMBLY
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
; STREET: 5 Palo Alto Square, 3000 El Camino Real
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/105,904B
; FILING DATE: 31-AUG-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057,594
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Neeley, Richard L.
; REGISTRATION NUMBER: 30,092
; REFERENCE/DOCKET NUMBER: ABIC-001/02US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)843-5000
; TELEFAX: (415)857-0663
; TELEX: 380846 Coolypa
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: AOA-Glu
US-08-105-904B-10

Query Match 100.0%; Score 11; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLOK 11
 Db 12 QPLALEGSLOK 22

RESULT 2
 US-08-114-877A-10
 ; Sequence 10, Application US/08114877A
 ; Patent No. 6174530
 ; GENERAL INFORMATION:
 ; APPLICANT: Rose, Keith
 ; TITLE OF INVENTION: HOMOGENOUS POLYOXIME COMPOSITIONS AND THEIR
 ; TITLE OF INVENTION: PREPARATION BY PARALLEL ASSEMBLY
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Cooley Godward Castro Huddleson & Tatum
 ; STREET: 5 Palo Alto Square
 ; CITY: Palo Alto
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 94036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/114,877A
 ; FILING DATE: 31-AUG-1993
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/057,594
 ; FILING DATE: 05-MAY-1993
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Neeley, Richard L.
 ; REGISTRATION NUMBER: 30,092
 ; REFERENCE/DOCKET NUMBER: ABIC-001/01US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 843 5070
 ; TELEFAX: (415) 857-0663
 ; TELEX: 380816 COOLEYPA
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 23 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHEetical: NO
 ; FEATURE:
 ; NAME/KEY: Modified-site
 ; LOCATION: 1
 ; OTHER INFORMATION: ROA-Glu
 ; US-08-114-877A-10

Query Match 100.0%; Score 11; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLOK 11
 Db 12 QPLALEGSLOK 22

RESULT 3
 US-08-537-928-10
 ; Sequence 10, Application US/08537928

Patent No. 6217873
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; APPLICANT: Offord, Robin E
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GRFN-001/03US
 ; CURRENT APPLICATION NUMBER: US/08/537,928
 ; EARLIER FILING DATE: 1996-01-05
 ; EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; EARLIER FILING DATE: 1994-05-05
 ; EARLIER APPLICATION NUMBER: 08/105,904
 ; EARLIER FILING DATE: 1993-08-31
 ; EARLIER APPLICATION NUMBER: 08/114,877
 ; EARLIER FILING DATE: 1993-08-31
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 ; US-08-537-928-10

Query Match 100.0%; Score 11; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLOK 11
 Db 12 QPLALEGSLOK 22

RESULT 4
 US-09-633-269-10
 ; Sequence 10, Application US/09633269
 ; Patent No. 6663869
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; APPLICANT: Offord, Robin E
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GRFN-001/03US
 ; CURRENT APPLICATION NUMBER: US/09/633,269
 ; CURRENT FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/08/537,928
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1996-01-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1994-05-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/105,904
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1993-08-31
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/114,877
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 ; US-09-633-269-10

Query Match 100.0%; Score 11; DB 4; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLOK 11
 Db 12 QPLALEGSLOK 22

RESULT 5
 US-09-633-269-10
 ; Sequence 10, Application US/09633269
 ; Patent No. 6663869
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; APPLICANT: Offord, Robin E
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GRFN-001/03US
 ; CURRENT APPLICATION NUMBER: US/09/633,269
 ; CURRENT FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/08/537,928
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1996-01-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1994-05-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/105,904
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1993-08-31
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/114,877
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 ; US-09-633-269-10

Query Match 100.0%; Score 11; DB 4; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLOK 11
 Db 12 QPLALEGSLOK 22

RESULT 6
 US-09-633-269-10
 ; Sequence 10, Application US/09633269
 ; Patent No. 6663869
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; APPLICANT: Offord, Robin E
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GRFN-001/03US
 ; CURRENT APPLICATION NUMBER: US/09/633,269
 ; CURRENT FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/08/537,928
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1996-01-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1994-05-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/105,904
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1993-08-31
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/114,877
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 ; US-09-633-269-10

Query Match 100.0%; Score 11; DB 4; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLOK 11
 Db 12 QPLALEGSLOK 22

RESULT 7
 US-09-633-269-10
 ; Sequence 10, Application US/09633269
 ; Patent No. 6663869
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; APPLICANT: Offord, Robin E
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GRFN-001/03US
 ; CURRENT APPLICATION NUMBER: US/09/633,269
 ; CURRENT FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/08/537,928
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1996-01-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1994-05-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/105,904
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1993-08-31
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/114,877
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 ; US-09-633-269-10

Query Match 100.0%; Score 11; DB 4; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLOK 11
 Db 12 QPLALEGSLOK 22

RESULT 8
 US-09-633-269-10
 ; Sequence 10, Application US/09633269
 ; Patent No. 6663869
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; APPLICANT: Offord, Robin E
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GRFN-001/03US
 ; CURRENT APPLICATION NUMBER: US/09/633,269
 ; CURRENT FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/08/537,928
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1996-01-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1994-05-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/105,904
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1993-08-31
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/114,877
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 ; US-09-633-269-10

Query Match 100.0%; Score 11; DB 4; Length 23;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLOK 11
 Db 12 QPLALEGSLOK 22

RESULT 9
 US-09-633-269-10
 ; Sequence 10, Application US/09633269
 ; Patent No. 6663869
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; APPLICANT: Offord, Robin E
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GRFN-001/03US
 ; CURRENT APPLICATION NUMBER: US/09/633,269
 ; CURRENT FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/08/537,928
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1996-01-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1994-05-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/105,904
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1993-08-31
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/114,877
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 ; US-09-633-269-10

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 06:03:19 ; Search time 38.9313 Seconds
(without alignments)
101.620 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLALEGSLQK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 3568729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: A_Geneseq_23Sep04:*
 - 2: Genesep1990s:*
 - 3: Genesep1990s:*
 - 4: Genesep2000s:*
 - 5: Genesep2001s:*
 - 6: Genesep2002s:*
 - 7: Genesep2003as:*
 - 8: Genesep2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	12	AAW06808	Single ch
2	53	100.0	18	ADI29027	Proinsuli
3	53	100.0	23	AAR60842	AOA-PePC
4	53	100.0	27	AAW81858	Human pro
5	53	100.0	29	AAB70847	Human pro
6	53	100.0	29	ADM34533	Preproins
7	53	100.0	32	AAP50011	Sequence
8	53	100.0	33	AAR83840	Part of 1
9	53	100.0	35	AAP40830	Sequence
10	53	100.0	35	AAR68897	Human ins
11	53	100.0	35	AAV08003	Human ins
12	53	100.0	35	AAO17829	Human ins
13	53	100.0	35	AA65678	Human pro
14	53	100.0	54	AA65687	Fragment
15	53	100.0	63	AAW06807	Single ch
16	53	100.0	64	AAR38432	Split (64)
17	53	100.0	64	AAR38433	Split (64)
18	53	100.0	65	AAO7660	Human ins
19	53	100.0	65	AAW94587	Insulin c
20	53	100.0	66	AAW06809	Met-Arg-M
21	53	100.0	72	AAU02065	Synthetic
22	53	100.0	83	AAE70639	Sequence
23	53	100.0	83	ADQ91446	Amino aci
24	53	100.0	84	AAE60248	Human pro
25	53	100.0	85	AAR20711	Ser A13,

26	53	100.0	85	7	ADFI7074	Human alb
27	53	100.0	86	1	AAF40829	Sequence
28	53	100.0	86	2	AAR20710	Proinsuli
29	53	100.0	86	2	AAR20699	Modified
30	53	100.0	86	2	AAR22327	Modified
31	53	100.0	86	2	AAR22330	Modified
32	53	100.0	86	2	AAR22331	Modified
33	53	100.0	86	2	AAR22332	Proinsuli
34	53	100.0	86	2	AAR20694	Modified
35	53	100.0	86	2	AAR22328	Modified
36	53	100.0	86	2	AAR20707	Modified
37	53	100.0	86	2	AAR22326	Modified
38	53	100.0	86	2	AAR22336	Modified
39	53	100.0	86	2	AAR20705	Proinsuli
40	53	100.0	86	2	AAR22329	Modified
41	53	100.0	86	2	AAR20693	Modified
42	53	100.0	86	2	AAR20712	Modified
43	53	100.0	86	2	AAR20704	Modified
44	53	100.0	86	2	AAR20692	Proinsuli
45	53	100.0	86	2	AAR22334	Proinsuli

ALIGNMENTS

RESULT 1
AAW06808
ID AAW06808 standard; peptide; 12 AA.
XX AC AAW06808;
XX AC
DT 16-MAR-1997 (first entry)
XX AC
DE Single chain insulin polypeptide bridging peptide.
XX AC
KW Single chain insulin; diabetes mellitus; gene therapy.
XX OS
XX OS
XX EP741188-A2.
XX PD 06-NOV-1996.
XX PF 03-MAY-1996; 96EP-00303133.
XX PR 05-MAY-1995; 95US-00435503.
XX PR 05-MAY-1995; 95US-00435762.
XX PA (ELIL) LILLY & CO ELI.
XX PI Chance RE, Dimarchi RD, Hoffmann JA, Long HB, Miller AR;
XX DR WPI; 1996-487391/49.
XX PT Single chain insulin polypeptide(s) - used for treating diabetes.
XX PS Disclosure; Page 16; 22pp; English.
XX CC Bridging peptides (AAW06808 and AAW00872) are utilised in novel single chain peptides of formula b-BP-a (see also AAW06806-07), where b is the insulin B-chain or its analogue, BP is a bridging peptide, and a is the insulin A-chain or its analogue. These peptides show high insulin bioactivity. Suitable bridging peptides should not contain sequences that interact with the IGF-1 receptor, nor should they contain cleavage sites. They are preferably derived from the C-peptide of human proinsulin so as to minimise potential immunogenicity, and most preferably consist of amino acids 54-65 of the C-peptide
XX SQ Sequence 12 AA;

Query Match 100.0%; Score 53; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. NO. 0.00059;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
 |||||
 Db 6 QPLALEGSLQK 16

RESULT 3
 AAR60842
 ID AAR60842 standard; peptide; 23 AA.
 XX
 AC AAR60842;
 XX
 DT 25-MAR-2003 (revised)
 DT 05-JUN-1995 (first entry)
 XX
 DE RoA-PepC COSM polyoxime component.
 XX
 DE MHC class II compound; MHC class II component; MHC class II alpha chain;
 KW MHC class II beta chain; peptide binding groove; spacerholder molecule;
 KW effector component; immune response; immune disorder; virucide;
 KW antibacterial; antiparasitic; cytostatic; immunosuppressive;
 KW gene therapy; viral infections; bacterial infection; parasitic infection;
 KW neoplastic disease; autoimmunity; toxicity.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO2004007528-A2.
 XX
 PD 22-JAN-2004.
 XX
 PF 11-JUL-2003; 2003WO-US021767.
 XX
 PR 12-JUL-2002; 2002US-0395494P.
 PR 22-JUL-2002; 2002US-0397593P.
 XX
 PA (DAND) DAMA FARBER CANCER INST INC.
 XX
 PI Wucherfenning KW, Seth N;
 XX
 DR WPI; 2004-122876/12.

Key Location/Qualifiers
 Modified-site 1
 FT /label= OTHER
 FT /note= "aminoxyacetyl-glutamic acid"
 XX
 PN WO9425071-A1.
 XX
 PD 10-NOV-1994.
 XX
 PF 05-MAY-1994; 94WO-IB000093.
 XX
 PR 05-MAY-1993; 93US-00057594.
 PR 31-AUG-1993; 93US-00105904.
 PR 31-AUG-1993; 93US-00114877.
 XX
 PA (ROSE/) ROSE K.
 PA (OFFO/) OFFORD R E.
 XX
 PI Rose K, Offord RE;
 XX
 DR WPI; 1994-357918/44.
 XX

Homo- and hetero-polyoxime compounds and their preparation - used for peptide presentation to antibodies and in cell imaging etc.
 PT
 PS Disclosure; Page 53; 85pp; English.
 XX
 CC Peptides given in AAR60833-62 are used as baseplates and COSMs for the preparation of polyoximes having varying spacing, charge, lipophilicity, CC valency, conformational restraints, solubility and other physical and biological properties. The RoA PepC COSM AAR60842 is based on the human CC proinsulin C peptide sequence. (Updated on 25-MAR-2003 to correct PN CC field.)
 XX
 SQ Sequence 23 AA;

Query Match 100.0%; Score 53; DB 2; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.0012;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
 |||||
 Db 12 QPLALEGSLQK 22

RESULT 4
 AAW81858
 ID AAW81858 standard; peptide; 27 AA.
 XX

QY 1 QPLALEGSLQK 11
 |||||
 Db 1 QPLALEGSLQK 11

RESULT 2
 ADI29027
 ID ADI29027 standard; peptide; 18 AA.
 XX
 AC ADI29027;
 XX
 DT 15-APR-2004 (first entry)
 XX
 DE Proinsulin (73-90) peptide SEQ ID NO:23.
 XX
 DE MHC class II compound; MHC class II component; MHC class II alpha chain;
 KW MHC class II beta chain; peptide binding groove; spacerholder molecule;
 KW effector component; immune response; immune disorder; virucide;
 KW antibacterial; antiparasitic; cytostatic; immunosuppressive;
 KW gene therapy; viral infections; bacterial infection; parasitic infection;
 KW neoplastic disease; autoimmunity; toxicity.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO2004007528-A2.
 XX
 PD 22-JAN-2004.
 XX
 PF 11-JUL-2003; 2003WO-US021767.
 XX
 PR 12-JUL-2002; 2002US-0395494P.
 PR 22-JUL-2002; 2002US-0397593P.
 XX
 PA (DAND) DAMA FARBER CANCER INST INC.
 XX
 PI Wucherfenning KW, Seth N;
 XX
 DR WPI; 2004-122876/12.

New MHC class II compound, useful for preparing a composition for treating immune disorders e.g. viral infections, bacterial infections, parasitic infections, neoplastic disease, autoimmunity or toxicity.
 PT
 PS Example 4; SEQ ID NO 23; 92pp; English.
 XX

The present invention describes an isolated MHC class II compound (I) comprising: (a) an MHC class II component comprising at least a portion of an MHC class II alpha chain and at least a portion of an MHC class II beta chain, such that the MHC class II alpha chain and MHC class II beta chain form a peptide binding groove; (b) a spacerholder molecule; and (c) an effector component, where the effector component is linked to the MHC class II component. Also described: (1) a pharmaceutical composition comprising the MHC class II molecule and a carrier; (2) a method of producing an MHC class II compound; (3) a method of directly identifying an antigen-specific T cell; (4) a method of regulating an immune response in a subject; (5) a method of treating an immune disorder in a subject; and (6) a method of regulating an immune response ex vivo in a subject; and (7) a method of treating an immune disorder ex vivo in a subject. (I) has virucide, antibacterial, antiparasitic, cytostatic and immunosuppressive activities, and can be used in gene therapy. The MHC class II compound (I) can be used for preparing a composition for treating immune disorders, e.g., viral infections, bacterial infections, parasitic infections, neoplastic disease, autoimmunity or toxicity. The present sequence represents a peptide which can be used to generate peptide-specific antibodies, which is used in an example from the present invention.

Query Match 100.0%; Score 53; DB 8; Length 18;
 Best Local Similarity 100.0%; Pred. No. 0.00094;

RESULT 18 AA;
 Query Match 100.0%; Score 53; DB 8; Length 18;
 Best Local Similarity 100.0%; Pred. No. 0.00094;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 11:39:31 ; Search time 21.7349 Seconds
(without alignments)
164.085 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 11
Sequence: 1 QPLALEGSLOK 11

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1370721 seqs, 324215800 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1370721

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	100.0	18	17	US-10-617-568-23
2	11	100.0	33	17	US-10-641-834-24
3	11	100.0	35	9	US-09-815-229-4
4	11	100.0	35	9	US-09-947-563-3
5	11	100.0	35	16	US-10-740-098-4
6	11	100.0	54	9	US-09-815-229-13
7	11	100.0	54	16	US-10-740-098-13
8	11	100.0	86	9	US-09-878-380-1
9	11	100.0	86	10	US-09-858-935B-4
10	11	100.0	86	13	US-10-028-410-2
11	11	100.0	86	13	US-10-054-873-4
12	11	100.0	86	14	US-10-444-326-2
13	11	100.0	86	14	US-10-271-869-4

Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 4, Appli
Sequence 5, Appli
Sequence 125, App
Sequence 3, Appli
Sequence 9, Appli
Sequence 6, Appli
Sequence 125, App
Sequence 125, App
Sequence 1, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 44, Appl
Sequence 44, Appl
Sequence 44, Appl
Sequence 44, Appl
Sequence 20, Appl
Sequence 44, Appl
Sequence 44, Appl
Sequence 44, Appl
Sequence 44, Appl
Sequence 3, Appli
Sequence 44, Appl
Sequence 63, Appl
Sequence 62, Appl
Sequence 1, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 15, Appl

86 15 US-10-444-262-2
86 15 US-10-444-649-2
86 15 US-10-444-701-2
96 9 US-09-947-563-4
96 9 US-09-947-563-5
110 9 US-09-205-658-125
110 9 US-09-815-229-3
110 9 US-09-804-409A-9
110 10 US-09-963-748C-6
110 10 US-09-963-693-125
110 14 US-10-038-686-1
110 14 US-10-328-813-2
110 14 US-10-383-285-2
110 14 US-10-345-563-2
110 14 US-10-321-717-2
110 15 US-10-411-037-44
110 15 US-10-411-026-44
110 15 US-10-410-962-44
110 15 US-10-411-049-44
110 15 US-10-700-725-20
110 16 US-10-410-930-44
110 16 US-10-410-997-44
110 16 US-10-411-012-44
110 16 US-10-287-994-44
110 16 US-10-740-098-3
110 16 US-10-410-913-44
117 9 US-09-280-030-63
130 9 US-09-269-439-1
31 9 US-09-878-380-2
31 9 US-09-878-380-2
31 15 US-10-430-752A-1
31 15 US-10-430-752A-15

ALIGNMENTS

RESULT 1
US-10-617-568-23
; Sequence 23, Application US/10617568
; Publication No. US20040197862A1
; GENERAL INFORMATION:
; APPLICANT: Dana-Farber Cancer Institute, Inc. et al.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: GENERATION OF MHC CLASS II COMPOUNDS BY
; TITLE OF INVENTION: PEPTIDE EXCHANGE
; FILE REFERENCE: DEN-044
; CURRENT APPLICATION NUMBER: US/10/617,568
; CURRENT FILING DATE: 2003-07-11
; PRIOR APPLICATION NUMBER: 60/395494
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 60/397893
; PRIOR FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-617-568-23

Query Match 100.0%; Score 11; DB 17; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.00037;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 QPLALEGSLOK 11
Db 6 QPLALEGSLOK 16
RESULT 2
US-10-641-834-24
; Sequence 24, Application US/10641834

```

; Publication No. US20040209830A1
; GENERAL INFORMATION:
; APPLICANT: Russell, Stephen James
; APPLICANT: Peng, Kah Whye
; TITLE OF INVENTION: SYSTEM FOR MONITORING THE EXPRESSION OF
; FILE OF INVENTION: FRANGENES
; FILE REFERENCE: 07039-292002
; CURRENT APPLICATION NUMBER: US/10/641,834
; CURRENT FILING DATE: 2003-05-18
; PRIOR APPLICATION NUMBER: 09/639,667
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: 60/149,168
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: C-peptide sequence
; US-10-641-834-24

```

```

Query Match 100.0%; Score 11; DB 17; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.00064;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 QPLALEGSLQK 11
Db 22 QPLALEGSLQK 32

```

```

RESULT 3
US-09-815-229-4
; Sequence 4, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS DISORDERS
; CURRENT APPLICATION NUMBER: US/09/815,229
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 4
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-815-229-4

```

```

Query Match 100.0%; Score 11; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.00067;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 QPLALEGSLQK 11
Db 24 QPLALEGSLQK 34

```

```

RESULT 4
US-09-947-563-3
; Sequence 3, Application US/09947563
; Patent No. US20020156234A1
; GENERAL INFORMATION:
; APPLICANT: Rubrodez, Franz-Josef
; APPLICANT: Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; insulin precursors having correctly bonded cystine bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farrah, Garrett &

```

```

; Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/947,563
; APPLICATION NUMBER: 09/134,836
; FILING DATE: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/134,836
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Leslie McDonnell
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02481.1600-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..35
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-947-563-3

```

```

Query Match 100.0%; Score 11; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.00067;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 QPLALEGSLQK 11
Db 24 QPLALEGSLQK 34

```

```

RESULT 5
US-10-740-098-4
; Sequence 4, Application US/10740098
; Publication No. US20040138101A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS DISORDERS
; CURRENT APPLICATION NUMBER: US/10/740,098
; FILE REFERENCE: P1786R1US
; CURRENT FILING DATE: 2003-12-17
; PRIOR APPLICATION NUMBER: US/09/815,229
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 4
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-740-098-4

```

```

Query Match 100.0%; Score 11; DB 16; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.00067;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 16:28:53 ; Search time 721.759 Seconds
(without alignments)
720.721 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 11
Sequence: 1 QPALLEGSLQK 11

Scoring table: OLIGO

Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9041963

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODE=frame+p2n model -DEV=xlp
-O=/cgr2.1/USPTO_sool_p/US10783095/runat_26102004_185454_25032/app_query.fasta_1.796
-DB=GenEmbl -QFWT=fastp -SUFFIX=oli.rge -MINMATCH=0.1 -LOOPCU=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=humar40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10783095@CCN 1.6148 @runat_26102004_185454_25032 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl.*

- 1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Table with 4 columns: Result No., Score, Match Length, DB ID, Description. Contains 4 rows of search results.

Table with 4 columns: Index, Score, Match Length, Description. Lists various sequences and their alignment scores.

ALIGNMENTS

RESULT 1
AR077834
LOCUS AR077834 68 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 18 from patent US 5962267.
ACCESSION AR077834
VERSION AR077834.1 GI:10004580
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68)
AUTHORS Shin,H.-C., Chang,S.-G., Kim,D.-Y. and Kim,C.-S.
TITLE Proinsulin derivative and process for producing human insulin
JOURNAL Patent: US 5962267-A 18 05-OCT-1999;
FEATURES
source 1..68
/organism="unknown"
/mol_type="unassigned DNA"

Table with 2 columns: Alignment Scores, Length. Shows statistical data for the alignment.

US-10-783-095-9 (1-11) x AR077834 (1-68)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
|||
Db 7 CAGCCGTTGGCGCTGGAAGGTTCCCTGCAGAAA 39

RESULT 2
AX107018 87 bp DNA linear PAT 30-APR-2001
LOCUS
DEFINITION Sequence 1 from Patent WO0125278.
ACCESSION AX107018
VERSION AX107018.1 GI:13922551
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.

REFERENCE 1
AUTHORS Habermann,P., Ertl,J., Meiwes,J. and Seipke,G.
TITLE C-peptide for the improved production of insulin and insulin analogues
JOURNAL Patent: WO 0125278-A 1 12-APR-2001;
Aventis Pharma Deutschland GmbH (DE)
FEATURES Location/Qualifiers
source 1..87
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="DNA kodierend fur C-PEPTIDEVARIANTE"

ORIGIN
Alignment Scores:
Pred. No.: 0.000329 Length: 87
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-783-095-9 (1-11) x AX107018 (1-87)
QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
|||
Db 52 CAGCCGTTGGCGCTGGAAGGTTCCCTGCAGAA 84

RESULT 3
XXHSIN
LOCUS
DEFINITION Synthetic gene for human insulin C-chain.
ACCESSION V01461 J02546
VERSION V01461.1 GI:58303
KEYWORDS artificial gene; gene synthesis.
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 105)
AUTHORS Hsiung,H.M., Sung,W.L., Brousseau,R., Wu,R. and Narang,S.A.
TITLE Synthesis of the human insulin gene. Part III. Chemical synthesis of 5'-phosphonoester group containing deoxyribonucleotides by the modified phosphorite method. Its application in the synthesis of seventeen fragments constituting human insulin C-chain DNA
JOURNAL Nucleic Acids Res. 8 (23), 5753-5765 (1980)
MEDLINE 81124297
PUBMED 7008029
FEATURES Location/Qualifiers
source 1..105
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
<1..>105
/note="(amino acids 31 to 65)"
/codon_start=1
/transl_table=11

CDS
ORIGIN
Alignment Scores:
Pred. No.: 0.000721 Length: 216
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-783-095-9 (1-11) x AX179492 (1-216)
QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
|||
Db 163 CACCCGTTGGCGCTGGAAGGTTCCCTGCAGAA 195

RESULT 5
I01821
LOCUS
DEFINITION Sequence 6 from Patent US 4801536.
ACCESSION I01821
VERSION I01821.1 GI:269797
KEYWORDS
SOURCE Unknown.

/product="insulin C chain"
/protein_id="CAA24707.1"
/db_xref="GI:929915"
/translation="FREADLQVGVVQLGGFGAGSLQPLALEGSLQKR"

ORIGIN
Alignment Scores:
Pred. No.: 0.000387 Length: 105
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-10-783-095-9 (1-11) x XXHSIN (1-105)
QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
|||
Db 70 CAACCTTGGCTCGAGGATCACTTCAAAG 102

RESULT 4
AX179492 216 bp DNA linear PAT 31-AUG-2001
LOCUS
DEFINITION Sequence 87 from Patent WO0131037.
ACCESSION AX179492
VERSION AX179492.1 GI:14599120
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.

REFERENCE 1
AUTHORS Ben-Nun,A., Kerlero de Rosbo,N. and Sappier,G.P.
TITLE Synthetic human genes and polypeptides and their use in the treatment of autoimmune diseases
JOURNAL Patent: WO 0131037-A 87 03-MAY-2001;
YEDA RESEARCH AND DEVELOPMENT CO. LTD. (IL)
FEATURES Location/Qualifiers
1..216
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
1..207
/note="unnamed protein product; synthetic construct"
/codon_start=1
/transl_table=11
/protein_id="CAC43900.1"
/db_xref="GI:14599121"
/translation="MPRELLPALLALWGPDPAAASHLVEALYLVSGRFFYTPKTR
TEAEAGAGSLQPLALEGSLQIKPR"

ORIGIN
Alignment Scores:
Pred. No.: 0.000721 Length: 216
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-783-095-9 (1-11) x AX179492 (1-216)
QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
|||
Db 163 CACCCGTTGGCGCTGGAAGGTTCCCTGCAGAA 195

RESULT 5
I01821
LOCUS
DEFINITION Sequence 6 from Patent US 4801536.
ACCESSION I01821
VERSION I01821.1 GI:269797
KEYWORDS
SOURCE Unknown.

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 28, 2004, 16:24:37 ; Search time 143.265 Seconds
(without alignments)
403.055 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 11
Sequence: 1 QPLALEGSLQK 11

Scoring table:
OLIGO 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1
Total number of hits satisfying chosen parameters: 8253336

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlp
-Q/cgn2_1/USPTO_spool_p/US10783095/runat_26102004_185454_25024/app_query.fasta_1.796
-DB=N Geneseg 23Sep04 -QFMT=fastap -SUFFIX=oli.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=oto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095@CN_1_1_1241@runat_26102004_185454_25024 -NCFU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMECUP=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOF=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N Geneseg_23Sep04:*
1: Genesegm1980s:*
2: Genesegm1990s:*
3: Genesegm2000s:*
4: Genesegm2001as:*
5: Genesegm2001bs:*
6: Genesegm2002as:*
7: Genesegm2002bs:*
8: Genesegm2003as:*
9: Genesegm2003bs:*
10: Genesegm2003cs:*
11: Genesegm2003ds:*
12: Genesegm2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Table with 6 columns: Result No., Score, Query Match, Length, DB ID, Description. Contains 6 rows of search results.

Table with 6 columns: ID, AX86082 standard; DNA; 66 BP., AC, AAX86082;, DT, 15-SEP-1999 (first entry), DE, PCR primer used to produce chimeric genes of the invention., XX, Chimeric fusion protein; immunodominant epitope; glutamate decarboxylase; GAD; insulin; insulin chain B; human T cell response; GAD65 peptide; Type I diabetes; insulin dependent diabetes mellitus; IDDM; Stiff Man syndrome; transplant; pancreatic cell; PCR primer; ss. OS Synthetic. OS Homo sapiens. XX MO9932136-A1. XX 01-JUL-1999. XX 23-DEC-1998; 98WO-US027408. XX 23-DEC-1997; 97US-0068648P. (ALEX-) ALEXION PHARM INC. (MUEL/) MUELLER J P. (MATI/) MATIS L A.

ALIGNMENTS

RESULT 1

Table with 6 columns: ID, AX86082 standard; DNA; 66 BP., AC, AAX86082;, DT, 15-SEP-1999 (first entry), DE, PCR primer used to produce chimeric genes of the invention., XX, Chimeric fusion protein; immunodominant epitope; glutamate decarboxylase; GAD; insulin; insulin chain B; human T cell response; GAD65 peptide; Type I diabetes; insulin dependent diabetes mellitus; IDDM; Stiff Man syndrome; transplant; pancreatic cell; PCR primer; ss. OS Synthetic. OS Homo sapiens. XX MO9932136-A1. XX 01-JUL-1999. XX 23-DEC-1998; 98WO-US027408. XX 23-DEC-1997; 97US-0068648P. (ALEX-) ALEXION PHARM INC. (MUEL/) MUELLER J P. (MATI/) MATIS L A.

PI Mueller JP, Matis LA, Wang Y;
XX WPI; 1999-405113/34.

XX Chimeric fusion proteins comprising immunodominant epitopes of glutamate
PT decarboxylase and insulin, used to treat patients that have or are
PT predicted to be at risk of developing Type I diabetes.

XX Example; Page 62; 67pp; English.

XX The specification describes chimeric fusion proteins comprising
CC immunodominant epitopes of glutamate decarboxylase (GAD) and insulin. The
CC chimeric fusion proteins comprises insulin chain B and individual peptide
CC moieties consisting of at least one GAD65 peptide capable of eliciting a
CC human T cell response, where the insulin chain B and GAD65 peptide are
CC covalently linked and the chimeric fusion protein is capable of eliciting
CC a human T cell response to insulin chain B and to each of the at least
CC one GAD65 peptides. The chimeric fusion proteins are used to treat
CC patients that are predicted to be at risk of developing type I diabetes
CC and those suffering from type I diabetes. The fusion proteins can also be
CC used to diagnose patients suffering from insulin dependent diabetes
CC mellitus (IDDM) and/or Stiff Man syndrome. This is also useful for
CC determining the suitability of patients as recipients of transplants of
CC pancreatic cells or tissues. The present sequence represents a primer
CC used to construct genes encoding the chimeric fusion proteins of the
CC invention

XX SQ Sequence 66 BP; 10 A; 27 C; 21 G; 8 T; 0 U; 0 Other;

Alignment Scores: Length: 66
Pred. No.: 0.00155 Matches: 11
Score: 11.00 Conservat: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 2 Gaps: 0
DB:

US-10-783-095-9 (1-11) x AAX86082 (1-66)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
|||
42 CAGCCCTTGGCCCTGGAGGGTCCCTGCAGAAG 10

RESULT 2
AAF61522
ID AAF61522 standard; DNA; 87 BP.

XX AC AAF61522;
XX DT 04-JUL-2001 (first entry)
XX DE Human proinsulin C-peptide encoding DNA.

XX KW Insulin; proinsulin; C-peptide; human; B-chain; A-chain; diabetes; ds.

XX OS Homo sapiens.

XX PN DE19947456-A1.

XX PD 05-APR-2001.

XX PF 02-OCT-1999; 99DE-01047456.

XX PR 02-OCT-1999; 99DE-01047456.

XX PA (AVET) AVENTIS PHARMA DEUT GMBH.

XX PI Habermann P, Ertl J, Meives J, Seipke G;

XX DR WPI; 2001-267042/28.

XX DR P-PSDB; AAB70847.

XX PT New synthetic derivatives of the C-peptide of proinsulin, useful in the

PT preparation of human insulin or insulin analogs in high yield.
XX Disclosure; Page 7; 12pp; German.

XX This invention describes novel synthetic derivatives (I) of the C-peptide
CC of proinsulin. Such novel derivatives of human insulin or insulin analogs
CC represented by (I) are of formula Fus-B(1-30)-RDVP-Yn-A(1-21). Fus =
CC optionally present fusion component of arbitrary sequence; B(1-30) = the
CC B-chain of human insulin; Y = amino acid chain having a basic amino acid
CC at the C-terminal; n = 2-50 (defining the chain length of Y), and A(1-21)
CC = the A-chain of human insulin; the A- and/or B-chain(s) may be modified
CC by amino acid exchanges, additions and/or deletions. (I) are
CC intermediates for human insulin or its analogs, which are used in the
CC treatment of diabetes. A claimed method of preparing human insulin (or
CC analogs) involves preparing (I), folding (I) so that disulfide bonds as
CC in human insulin can form, enzymatically removing the RDVP-Yn part and
CC Fus (if present) and purifying the product. Insulin (or analogs) can be
CC prepared in increased yields via (I). Specifically the expression yield
CC can be increased around 20% compared with that obtained using plasmid
CC pINR90 and almost 5-fold compared with that obtained using plasmid
CC pINR302d. The control of enzymatic processing is also improved. The
CC folding rate is comparable with that obtained using simian proinsulin
CC encoded by pINR90d. This sequence encodes the human proinsulin C-peptide
CC which is described in the method of the invention

XX SQ Sequence 87 BP; 10 A; 27 C; 37 G; 13 T; 0 U; 0 Other;

Alignment Scores: Length: 87
Pred. No.: 0.00201 Matches: 11
Score: 11.00 Conservat: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 4 Gaps: 0
DB:

US-10-783-095-9 (1-11) x AAF61522 (1-87)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
|||
52 CAGCCCTTGGCCCTGGAGGGTCCCTGCAGAAG 84

RESULT 3
AAF45977
ID AAF45977 standard; cDNA; 159 BP.

XX AC AAF45977;

XX DT 16-MAR-1997 (first entry)

XX DE Oligonucleotide for 3' end of single chain insulin.

XX KW Single chain insulin; diabetes mellitus; gene therapy; ss.

XX OS Synthetic.

XX PN EP741188-A2.

XX PD 06-NOV-1996.

XX PF 03-MAY-1996; 96EP-00303133.

XX PR 05-MAY-1995; 95US-00435503.

XX PR 05-MAY-1995; 95US-00435762.

XX PA (ELIL) LILLY & CO ELI.

XX PI Chance RE, Dimarchi RD, Hoffmann JA, Long HB, Miller AR;

XX DR WPI; 1996-487991/49.

XX PT Single chain insulin polypeptide(s) - used for treating diabetes.

XX PS Example 7; Page 15; 22pp; English.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 20:02:28 ; Search time 32.6024 Seconds
(without alignments)
239.819 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 11
Sequence: 1 QPLALEGSLQK 11

Scoring table:
OLIGO 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1641062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlp
-C/cgn2_1/USPTC spool_p/US10783095/runat_26102004_185455_25065/app_query.fasta_1.796
-DB=Issued patents.NA -OFMT=fastap -SUFFIX=oli.rni -MINMATCH=0.1 -LTOPCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THE SCORE=quality -THE MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=exc -HEAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095@cgn_1_1_213@runat_26102004_185455_25065 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

- Issued Patents.NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Table with 5 columns: Result No., Score, Match, Length, DB ID, Description. Contains 12 rows of search results.

Table with 5 columns: Hit No., Score, Match, Length, DB ID, Description. Contains 45 rows of search results.

ALIGNMENTS

RESULT 1

US-08-600-783-18
; Sequence 18, Application US/08600783
; Patent No. 5962267
; GENERAL INFORMATION:
; APPLICANT: SHIN, Hang Cheol
; APPLICANT: CHANG, Seung Gu
; APPLICANT: KIM, Dae Young
; APPLICANT: KIM, Chong Suhll
; TITLE OF INVENTION: Proinsulin Derivative and Process
; TITLE OF INVENTION: for Producing Human Insulin
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SHIN, Hang Cheol
; STREET: Ssangma-Hanshin Apt. 102-1206,
; CITY: Kwangmyung-shi
; STATE: Kyungki-do
; COUNTRY: Republic of Korea
; ZIP: 423-030
; ADDRESSEE: CHANG, Seung Gu
; STREET: Hyundai Apt. 71-203, Apxujong-dong,
; CITY: Kangnam-ku
; STATE: Seoul
; COUNTRY: Republic of Korea
; ZIP: 135-110
; ADDRESSEE: KIM, Dae Young
; STREET: Sosa Jukong Apt. 108-202, Sosa Bon-dong,
; CITY: Bucheon-shi
; STATE: Kyungki-do
; COUNTRY: Republic of Korea
; ZIP: 422-230
; ADDRESSEE: KIM, Chong Suhll
; STREET: Garden Heights Apt. 202-801, #100,

STREET: Hwangkeum-dong, Soosung-ku
 CITY: Taegu
 STATE: Taegu
 COUNTRY: Republic of Korea
 ZIP: 706-040
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy Disk, 3.5 inch, 1.44MB storage
 COMPUTER: IBM PC/AT
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Word Perfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/600,783
 FILING DATE:
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: KR 95-2751
 FILING DATE: 15-FEB-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Shahan Islam
 REGISTRATION NUMBER: 32,507
 REFERENCE/DOCKET NUMBER:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-1000
 TELEFAX: (212) 953-7249
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 68 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: oligonucleotide
 ANTI-SENSE: yes
 US-08-600-783-18

Alignment Scores:
 Pred. No.: 0.00207 Length: 68
 Score: 11.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-783-095-9 (1-11) x US-08-600-783-18 (1-68)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnlys 11
 |||||
 DB 7 CAGCCGTTGGCGCTGGAAGGTTCCCTGCAGAA 39

RESULT 2

US-09-676-787-1
 ; Sequence 1, Application US/09676787
 ; Patent No. 6534288
 ; GENERAL INFORMATION:
 ; APPLICANT: HABERMANN, Paul
 ; APPLICANT: BRILL, Johann
 ; APPLICANT: MEIWES, Johannes
 ; APPLICANT: SEIPEKE, Gerhard
 ; TITLE OF INVENTION: C PEPTIDE FOR IMPROVED PREPARATION OF INSULIN AND INSULIN ANALOGS
 ; FILE REFERENCE: 38005-0110
 ; CURRENT APPLICATION NUMBER: US/09/676,787
 ; CURRENT FILING DATE: 2000-10-02
 ; PRIOR APPLICATION NUMBER: DE 199 47 456.7
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 87
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-676-787-1

Alignment Scores:
 Pred. No.: 0.00262 Length: 87

Score: 11.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-783-095-9 (1-11) x US-09-676-787-1 (1-87)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnlys 11
 |||||
 DB 52 CAGCCGTTGGCGCTGGAAGGTTCCCTGCAGAA 84

RESULT 3

US-08-600-783-17
 ; Sequence 17, Application US/08600783
 ; Patent No. 5962267
 ; GENERAL INFORMATION:
 ; APPLICANT: SHIN, Hang Cheol
 ; APPLICANT: CHANG, Seung Gu
 ; APPLICANT: KIM, Dae Young
 ; APPLICANT: KIM, Chong Suh1
 ; TITLE OF INVENTION: Proinsulin Derivative and Process
 ; TITLE OF INVENTION: for Producing Human Insulin
 ; NUMBER OF SEQUENCES: 36
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SHIN, Hang Cheol
 ; STREET: Ssangma-Hanshin Apt. 102-1206,
 ; STREET: #245 Cholsan-dong
 ; CITY: Kwangmyung-shi
 ; STATE: Kyungki-do
 ; COUNTRY: Republic of Korea
 ; ZIP: 423-030
 ; ADDRESSEE: CHANG, Seung Gu
 ; STREET: Hyundai Apt. 71-203, Apkujong-dong,
 ; STREET: Kangnam-ku
 ; CITY: Seoul
 ; STATE: Seoul
 ; COUNTRY: Republic of Korea
 ; ZIP: 135-110
 ; ADDRESSEE: KIM, Dae Young
 ; STREET: Sosa Jukong Apt. 108-202, Sosa Bon-dong,
 ; STREET: Sosa-ku
 ; CITY: Bucheon-shi
 ; STATE: Kyungki-do
 ; COUNTRY: Republic of Korea
 ; ZIP: 422-230
 ; ADDRESSEE: KIM, Chong Suh1
 ; STREET: Garden Heights Apt. 202-801, #100,
 ; STREET: Hwangkeum-dong, Soosung-ku
 ; CITY: Taegu
 ; STATE: Taegu
 ; COUNTRY: Republic of Korea
 ; ZIP: 706-040
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy Disk, 3.5 inch, 1.44MB storage
 ; COMPUTER: IBM PC/AT
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Word Perfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/600,783
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: KR 95-2751
 ; FILING DATE: 15-FEB-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Shahan Islam
 ; REGISTRATION NUMBER: 32,507
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-1000
 ; TELEFAX: (212) 953-7249
 ; INFORMATION FOR SEQ ID NO: 17:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 22:07:29 ; Search time 580.217 Seconds
(without alignments)
97.211 Million cell updates/sec

Title: US-10-783-095-9

Perfect score: 11

Sequence: 1 QPLALESLQK 11

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3413475 seqs, 2563800928 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6814041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPCO_specolp/US10783095/runat_26102004_185456_25148/app_query.fasta_1.796
-DB=PublishedApplications NA -QFWT=fastap -SUFFIX=oli.rnpp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=cligo
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10783095@cgn1.1.1189@runat_26102004_185456_25148
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

†

Result No.	Score	Query Match	Length	DB	ID	Description
1	11	100.0	258	15	US-10-383-285-1	Sequence 1, Appli
2	11	100.0	260	10	US-09-807-742-16	Sequence 16, Appl
3	11	100.0	280	10	US-09-807-742-17	Sequence 17, Appl
4	11	100.0	260	17	US-10-746-149-23	Sequence 23, Appl
5	11	100.0	260	18	US-10-609-019-21	Sequence 21, Appl
6	11	100.0	330	15	US-10-346-563-1	Sequence 1, Appli
7	11	100.0	330	16	US-10-321-717-1	Sequence 1, Appli
8	11	100.0	351	9	US-09-280-030-49	Sequence 49, Appl
9	11	100.0	390	9	US-09-280-030-48	Sequence 48, Appl
10	11	100.0	450	9	US-09-804-409A-10	Sequence 10, Appl
11	11	100.0	450	10	US-09-969-748C-5	Sequence 5, Appli
12	11	100.0	450	15	US-10-328-813-1	Sequence 1, Appli
13	11	100.0	450	16	US-10-411-037-43	Sequence 43, Appl
14	11	100.0	450	16	US-10-411-026-43	Sequence 43, Appl
15	11	100.0	450	16	US-10-410-962-43	Sequence 43, Appl
16	11	100.0	450	16	US-10-411-049-43	Sequence 43, Appl
17	11	100.0	450	17	US-10-410-930-43	Sequence 43, Appl
18	11	100.0	450	17	US-10-410-997-43	Sequence 43, Appl
19	11	100.0	450	17	US-10-411-012-43	Sequence 43, Appl
20	11	100.0	450	17	US-10-287-994-43	Sequence 43, Appl
21	11	100.0	450	17	US-10-410-913-43	Sequence 43, Appl
22	11	100.0	1182	9	US-09-919-344-27	Sequence 27, Appl
23	11	100.0	1626	18	US-10-609-019-27	Sequence 27, Appl
24	11	100.0	2499	10	US-09-775-508A-1	Sequence 1, Appli
25	11	100.0	10487	17	US-10-746-149-44	Sequence 44, Appl
26	11	100.0	10487	18	US-10-609-019-32	Sequence 32, Appl
27	11	100.0	10512	17	US-10-746-149-42	Sequence 42, Appl
28	11	100.0	10512	18	US-10-609-019-31	Sequence 31, Appl
29	11	100.0	10880	17	US-10-746-149-47	Sequence 47, Appl
30	11	100.0	10895	18	US-10-609-019-41	Sequence 41, Appl
31	11	100.0	11255	17	US-10-746-149-43	Sequence 43, Appl
32	11	100.0	11271	18	US-10-609-019-42	Sequence 42, Appl
33	11	100.0	11332	18	US-10-609-019-43	Sequence 43, Appl
34	10	90.9	68	16	US-10-430-752A-9	Sequence 9, Appli
35	10	90.9	74	16	US-10-430-752A-7	Sequence 7, Appli
36	10	90.9	375	16	US-10-221-677-25	Sequence 25, Appl
37	10	90.9	1944	16	US-10-221-677-26	Sequence 26, Appl
38	10	90.9	4846	16	US-10-430-752A-14	Sequence 14, Appl
39	8	72.7	333	17	US-10-419-539-2	Sequence 2, Appli
40	8	72.7	598	13	US-10-013-032-11	Sequence 11, Appl
41	8	72.7	633	13	US-10-013-032-12	Sequence 12, Appl
42	8	72.7	666	13	US-10-013-032-13	Sequence 13, Appl
43	8	72.7	1680	17	US-10-419-539-3	Sequence 3, Appli
44	8	72.7	1217	17	US-10-419-539-4	Sequence 4, Appli
45	7	63.6	323	9	US-09-783-590-1388	Sequence 1388, Ap

ALIGNMENTS

RESULT 1

US-10-383-285-1 ; Sequence 1, Application US/10383285
; Publication No. US20030200556A1
; GENERAL INFORMATION:
; APPLICANT: STREATFIELD, STEPHEN
; APPLICANT: HOWARD, JOHN
; TITLE OF INVENTION: PRODUCTION OF INSULIN AND INSULIN-LIKE PROTEINS IN
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 10040
; CURRENT APPLICATION NUMBER: US/10/383,285
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: 60/362,874
; PRIOR FILING DATE: 2002-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 258
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-383-285-1

Alignment Scores: Length: 258
 Pred. No.: 0.00302
 Score: 11.00
 Matches: 11
 Conservative: 0
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Mismatches: 0
 Indels: 0
 Query Match: 100.00%
 DB: 15
 Gaps: 0

US-10-783-095-9 (1-11) x US-10-383-285-1 (1-258)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
 |||||
 Db 160 CAGCCCTGGCCCTCGAGGGCTCCCTCCAGAAG 192

RESULT 2
 US-09-807-742-16
 ; Sequence 16, Application US/09807742
 ; Publication No. US20030204864A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DANIELL, HENRY
 ; TITLE OF INVENTION: PRODUCTION OF PHARMACEUTICAL PROTEINS IN TRANSGENIC
 ; FILE OF INVENTION: PLASTIDS
 ; TITLE REFERENCE: 1465-PCT-US-00
 ; CURRENT APPLICATION NUMBER: US/09/807,742
 ; CURRENT FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: PCT/US01/06288
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 16
 ; LENGTH: 260
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-807-742-16

Alignment Scores: Length: 260
 Pred. No.: 0.00304
 Score: 11.00
 Matches: 11
 Conservative: 0
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Mismatches: 0
 Indels: 0
 Query Match: 100.00%
 DB: 10
 Gaps: 0

US-10-783-095-9 (1-11) x US-09-807-742-16 (1-260)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
 |||||
 Db 160 CAGCCCTGGCCCTCGAGGGCTCCCTCCAGAAG 192

RESULT 3
 US-09-807-742-17
 ; Sequence 17, Application US/09807742
 ; Publication No. US20030204864A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DANIELL, HENRY
 ; TITLE OF INVENTION: PRODUCTION OF PHARMACEUTICAL PROTEINS IN TRANSGENIC
 ; FILE OF INVENTION: PLASTIDS
 ; TITLE REFERENCE: 1465-PCT-US-00
 ; CURRENT APPLICATION NUMBER: US/09/807,742
 ; CURRENT FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: PCT/US01/06288
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 17
 ; LENGTH: 260
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Chloroplast
 US-09-807-742-17

Alignment Scores: Length: 260
 Pred. No.: 0.00304
 Score: 11.00
 Matches: 11
 Conservative: 0
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Mismatches: 0
 Indels: 0
 Query Match: 100.00%
 DB: 10
 Gaps: 0

US-10-783-095-9 (1-11) x US-09-807-742-17 (1-260)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
 |||||
 Db 160 CAACCTTAGCTTTAGAAGGTTCTTTACAAAAA 192

RESULT 4
 US-10-746-149-23
 ; Sequence 23, Application US/10746149
 ; Publication No. US20040172667A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cooper, Richard K.
 ; APPLICANT: Fioretti, William C.
 ; APPLICANT: Cadd, Gary G.
 ; TITLE OF INVENTION: Administration of Transposon-Based Vectors to Reproductive Organs
 ; FILE REFERENCE: 51687-0280 (51687-294923)
 ; CURRENT APPLICATION NUMBER: US/10/746,149
 ; CURRENT FILING DATE: 2003-12-24
 ; PRIOR APPLICATION NUMBER: US 60/392,415
 ; PRIOR FILING DATE: 2002-06-26
 ; PRIOR APPLICATION NUMBER: US 60/441,381
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,447
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,405
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,502
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,392
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,377
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 10/609,019
 ; PRIOR FILING DATE: 2003-06-26
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 23
 ; LENGTH: 260
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Synthetic
 US-10-746-149-23

Alignment Scores: Length: 260
 Pred. No.: 0.00304
 Score: 11.00
 Matches: 11
 Conservative: 0
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Mismatches: 0
 Indels: 0
 Query Match: 100.00%
 DB: 17
 Gaps: 0

US-10-783-095-9 (1-11) x US-10-746-149-23 (1-260)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
 |||||
 Db 160 CAGCCCTGGCCCTCGAGGGCTCCCTCCAGAAG 192

RESULT 5
 US-10-609-019-21
 ; Sequence 21, Application US/10609019
 ; Publication No. US20040197910A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cooper, Richard K.

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
 Run on: October 28, 2004, 11:13:10 ; Search time 5.6988 Seconds
 (without alignments)
 185.721 Million cell updates/sec

Title: US-10-783-095-9
 Perfect score: 11
 Sequence: 1 QPLALEGSLQK 11
 Scoring table: OLIGO
 Gapop 60.0 , Gapext 60.0

Searched: 283416 segs, 96216763 residues
 Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 79:*
 1: pir1:*
 2: pir2:*
 3: pir3:*
 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	11	100.0	110	1	insulin precursor
2	11	100.0	110	1	IPHU
3	11	100.0	110	2	insulin precursor
4	11	100.0	110	2	JQ0178
5	7	63.6	110	1	insulin precursor
6	7	63.6	456	2	IPDG
7	6	54.5	80	2	hypothetical prote
8	6	54.5	110	1	T40386
9	6	54.5	156	2	hypothetical prote
10	6	54.5	199	2	insulin precursor
11	6	54.5	205	2	IPGP
12	6	54.5	218	2	glutamine-fructose
13	6	54.5	220	2	hypothetical prote
14	6	54.5	230	2	H64230
15	6	54.5	232	2	SOJ protein (so)
16	6	54.5	265	2	hypothetical prote
17	6	54.5	267	2	photosystem II oxy
18	6	54.5	272	2	movement protein -
19	6	54.5	275	2	probable transpor
20	6	54.5	275	2	hypothetical prote
21	6	54.5	297	2	enoyl-lacyl-carrie
22	6	54.5	322	2	hypothetical prote
23	6	54.5	344	2	hypothetical prote
24	6	54.5	344	2	porphobilinogen de
25	6	54.5	358	2	O-sialoglycoprotei
26	6	54.5	383	2	probable Transposa
27	6	54.5	384	2	deoxyhypusine synt
28	6	54.5	405	2	hypothetical prote
29	6	54.5	440	2	probable PAD-depen
30	6	54.5	440	2	probable indole-3-

30	6	54.5	455	2	T00584
31	6	54.5	464	2	indole-3-acetate b
32	6	54.5	479	2	intermediate filam
33	6	54.5	479	2	hydrogenase-4 comp
34	6	54.5	479	2	hydrogenase 4 memb
35	6	54.5	484	2	hydrogenase 4 memb
36	6	54.5	484	2	thiazole biosynthe
37	6	54.5	495	2	NADH2 gehydrogenas
38	6	54.5	518	2	n-acetylglucosamin
39	6	54.5	524	2	ATP-dependent RNA
40	6	54.5	524	2	conserved hypochet
41	6	54.5	535	2	atp dependent hell
42	6	54.5	535	2	glutamine-fructose
43	6	54.5	597	2	glutamine-fructose
44	6	54.5	598	2	glutamine-fructose
45	6	54.5	608	2	glutamine-fructose
46	6	54.5	608	2	hypothetical prote

ALIGNMENTS

RESULT 1

insulin precursor [validated] - human
 N;Alternate names: preproinsulin
 C;Species: Homo sapiens (man)
 C;Date: 23-Oct-1981 #sequence, revision 23-Oct-1981 #text change 09-Jul-2004
 C;Accession: A93222; A94253; A93216; A94251; A93144; A92075; A91186; I58114; A01579; S58
 R;Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Cordell, B.; Tischer, E.; Goodman, H.M.
 Nature 284, 26-32, 1980
 A;Title: Sequence of the human insulin gene.
 A;Reference number: A93222; MUID:80120725; PMID:6243748
 A;Accession: A93222
 A;Molecule type: DNA
 A;Residues: 1-110 <BEL>
 A;Cross-references: UNIPROT:P01308; GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G3866828
 R;Ullrich, A.; Dull, T.J.; Gray, A.; Brosius, J.; Sures, I.
 Science 209, 612-615, 1980
 A;Title: Genetic variation in the human insulin gene.
 A;Reference number: A94253; MUID:80236313; PMID:6248962
 A;Accession: A94253
 A;Molecule type: DNA
 A;Residues: 1-110 <ULL>
 A;Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G3866828
 R;Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.
 Nature 282, 525-527, 1979
 A;Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.
 A;Reference number: A93216; MUID:80054779; PMID:503234
 A;Accession: A93216
 A;Molecule type: mRNA
 A;Residues: 1-110 <BEL2>
 A;Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G3866828
 R;Sures, I.; Goedel, D.V.; Gray, A.; Ullrich, A.
 Science 208, 57-59, 1980
 A;Title: Nucleotide sequence of human preproinsulin complementary DNA.
 A;Reference number: A94251; MUID:80147417; PMID:6927840
 A;Accession: A94251
 A;Molecule type: mRNA
 A;Residues: 1-110 <SUR>
 A;Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G3866828
 R;Nicol, D.S.H.W.; Smith, L.F.
 Nature 187, 483-485, 1960
 A;Title: Amino-acid sequence of human insulin.
 A;Reference number: A93144
 A;Accession: A93144
 A;Molecule type: protein
 A;Residues: 25-54;90-110 <NIC>
 R;Oyer, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.
 J. Biol. Chem. 246, 1375-1386, 1971
 A;Title: Studies on human proinsulin. Isolation and amino acid sequence of the human pan
 A;Reference number: A92075; MUID:71116410; PMID:5101771
 A;Accession: A92075
 A;Molecule type: protein

A;Residues: 57-87 <OYE>
 R;Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
 Eur. J. Biochem. 20, 190-199, 1971
 A;Title: Amino acid sequence of the C-peptide of human proinsulin.
 A;Reference number: A91186; MUID:71257722; PMID:5560404
 A;Accession: A91186
 A;Molecule type: protein
 A;Residues: 57-87 <OX>
 R;Lucassen, A.M.; Jullier, C.; Beressi, J.P.; Boitard, C.; Froguel, P.; Lathrop, M.; Bell
 Nature Genet. 4, 305-310, 1993
 A;Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment
 A;Reference number: I58114; MUID:93364428; PMID:8358440
 A;Accession: I58114
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-59,63-110 <RES>
 A;Cross-references: GB:L15440; NID:G307071; PIDN:AAA59179.1; PID:G307072
 R;Sieber, P.; Kamber, B.; Hartmann, A.; Joehli, A.; Riniker, B.; Rittel, W.
 Helv. Chim. Acta 57, 2617-2621, 1974
 A;Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindungen.
 A;Reference number: A91636; MUID:75077277; PMID:4443293
 A;Contents: annotation; synthesis
 A;Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was identical
 A;Note: article in German with English abstract
 R;Naithani, V.K.
 Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973
 A;Title: The synthesis of C-peptide of human proinsulin.
 A;Reference number: A91658; MUID:75040007; PMID:4803504
 A;Contents: annotation; synthesis of residues 57-87
 R;Geiger, R.; Jaeger, G.; Koenig, W.
 Chem. Ber. 106, 2347-2352, 1973
 A;Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9
 A;Reference number: A90914
 A;Contents: annotation; synthesis of residues 57-87
 R;Kaufmann, J.E.; Irmingier, J.C.; Halban, P.A.
 Biochem. J. 310, 869-874, 1995
 A;Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide juncti
 A;Reference number: S58661; MUID:96013185; PMID:7575420
 A;Contents: annotation; site-directed mutagenesis study of proteolytic processing
 C;Genetics:
 A;Gene: GDB:INS
 A;Cross-references: GDB:119349; OMIM:176730
 A;Map position: 11p15.3-11p15.5
 A;Introns: 63/1
 C;Superfamily: insulin
 C;Keywords: hormone; pancreas
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-54/Domain: insulin chain B #status experimental <BCH>
 F;25-54,90-110/Product: insulin #status experimental <MAT>
 F;57-87/Domain: connecting C peptide #status experimental <CPEP>
 F;90-110/Domain: insulin chain A #status experimental <ACH>
 F;31-96,43-109,95-100/Disulfide bonds: #status experimental
 Query Match 100.0%; Score 11; DB 1; Length 110;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 100.0%; Score 11; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
 |||||
 DB 78 QPLALEGSLQK 88

RESULT 2
 B42179
 insulin precursor - green monkey
 C;Species: Cercopithecus aethiops (green monkey, grivet)
 C;Date: 04-Mar-1993 #sequence revision 18-Nov-1994 #text_change 09-Jul-2004
 C;Accession: B42179; A05232; S16494; S22056
 R;Seino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of m
 A;Reference number: A42179; MUID:92219953; PMID:11560757
 A;Accession: B42179

QY 1 QPLALEGSLQK 11
 |||||
 DB 78 QPLALEGSLQK 88

RESULT 4
 A42179
 insulin precursor - chimpanzee
 C;Species: Pan troglodytes (chimpanzee)
 C;Date: 04-Mar-1993 #sequence revision 18-Nov-1994 #text_change 09-Jul-2004
 C;Accession: A42179; S22058
 R;Seino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of m
 A;Reference number: A42179; MUID:92219953; PMID:11560757
 A;Accession: A42179

QY 1 QPLALEGSLQK 11
 |||||
 DB 78 QPLALEGSLQK 88

RESULT 3
 JQ0178
 insulin precursor - crab-eating macaque
 C;Species: Macaca fascicularis (crab-eating macaque)
 C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
 C;Accession: JQ0178
 R;Wetkham, W.; Gronenberg, J.; Leineweber, M.; Wengenmayer, F.; Winnacker, E.L.
 Gene 19, 179-183, 1982
 A;Title: The nucleotide sequence of cDNA coding for proinsulin from the primate Macac
 A;Reference number: JQ0178; MUID:83080474; PMID:6184262
 A;Accession: JQ0178
 A;Molecule type: mRNA
 A;Residues: 1-110 <WET>
 A;Cross-references: UNIPROT:P30406; GB:J00336; NID:G342121; PIDN:AAA36849.1; PID:G342122
 C;Superfamily: insulin
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-54,90-110/Product: insulin #status predicted <MAT>
 F;25-54/Domain: insulin chain B #status predicted <BCH>
 F;55-89/Domain: insulin connecting C peptide #status predicted <CPT>
 F;90-110/Domain: insulin chain A #status predicted <ACH>
 F;31-96,43-109,95-100/Disulfide bonds: #status predicted
 Query Match 100.0%; Score 11; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
 |||||
 DB 78 QPLALEGSLQK 88

RESULT 4
 A42179
 insulin precursor - chimpanzee
 C;Species: Pan troglodytes (chimpanzee)
 C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C;Accession: A42179; S22058
 R;Seino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of m
 A;Reference number: A42179; MUID:92219953; PMID:11560757
 A;Accession: A42179

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 19:58:09 ; Search time 1249.89 Seconds
(without alignments)
320.697 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 11
Sequence: 1 QPLALEGSLQK 11

Scoring table: OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Word size: 1

Total number of hits satisfying chosen parameters: 65641631

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlp
-O=/cpn2.1/USPTO.spool.p/US10783095/runat_26102004.185454.25046/app_query.fasta_1.796
-DB=EST_QPWT=fastap -SUFFIX=oli.rst -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=40
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptb
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10783095 @CGN 1 1.9525 @runat_26102004.185454.25046 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Table with 5 columns: Result No., Score, Match, Length, DB ID, Description. Contains 9 rows of search results.

Table with 5 columns: C, 10, 11, 100.0, 174, 5, BQ272395. Contains 45 rows of sequence alignment data.

ALIGNMENTS

RESULT 1
LOCUS BQ267254/c
DEFINITION BQ267254.1 GI:20492319
ACCSSION BQ267254
VERSION BQ267254.1
KEYWORDS EST
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 118)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,J., Marra,M., Page,D., Wylie,T., Martin,J.G., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarisshvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE Unpublished (2000)
JOURNAL Endocrine Pancreas Consortium
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@hwh.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a

sequence.
BQ267254 118 bp mRNA linear EST 15-JUL-2003
ik01b07.y1 Human insulinoma Homo sapiens cdNA clone IMAGE:5779548
5', similar to SW:INS_HUMAN P01308 INSULIN PRECURSOR. [1]; MRNA

clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco.

FEATURES source

```
1..118
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5779548"
/tissue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Human insulinoma"
/note="Organ: pancreas; Vector: pBluescript SK-; Site 1: XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to pBluescript SK- by Dr. H. Inoue following the Washington University protocol (http://genome.wustl.edu/est/lambda_protocol.shtml). Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."
```

ORIGIN

Alignment Scores: 0.0897 Length: 118
Pred. No.: 11.00 Matches: 11
Score: 11.00
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-783-095-9 (1-11) x BQ267254 (1-118)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
Db 113 CAGCCCTTGGCCCTGGAGGGTCCCTGCAGAAG 81

RESULT 2
BQ272508 118 bp mRNA linear EST 15-JUL-2003
LOCUS ik01b07.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5779548
DEFINITION 3' similar to SWINS_HUMAN P01308 INSULIN PRECURSOR. [1] ; mRNA sequence.

ACCESSION BQ272508
VERSION BQ272508.1 GI:20497577
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 118)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,K., Cole,R., Tsagarishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Other ESTs: ik01b07.y1

Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)

clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Possible reversed clone: similarity on wrong strand
Seq primer: -40UP from Gibco.

FEATURES source

```
1..118
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5779548"
/tissue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Human insulinoma"
/note="Organ: pancreas; Vector: pBluescript SK-; Site 1: XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to pBluescript SK- by Dr. H. Inoue following the Washington University protocol (http://genome.wustl.edu/est/lambda_protocol.shtml). Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."
```

ORIGIN

Alignment Scores: 0.0897 Length: 118
Pred. No.: 11.00 Matches: 11
Score: 11.00
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-783-095-9 (1-11) x BQ272508 (1-118)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
Db 5 CAGCCCTTGGCCCTGGAGGGTCCCTGCAGAAG 37

RESULT 3
BU073408 118 bp mRNA linear EST 27-AUG-2002
LOCUS im35e06.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:6036971
DEFINITION 3' similar to SWINS_HUMAN P01308 INSULIN PRECURSOR. [1] ; mRNA sequence.

ACCESSION BU073408
VERSION BU073408.1 GI:22514597
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 118)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:57:30 ; Search time 27.8313 Seconds
(without alignments)
227.410 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 11
Sequence: 1 QPLALEGSLOK 11

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 1825181 seqs, 575374646 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : UniProt 02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	100.0	65	Q8hz80	pongo pygma
2	11	100.0	65	Q8hz81	gorilla gor
3	11	100.0	110	INS_CERRA	cercopithe
4	11	100.0	110	INS_HUMAN	homo sapien
5	11	100.0	110	INS_MACFA	macaca fasc
6	11	100.0	110	INS_PANTR	pan troglod
7	11	100.0	110	INS_PONPY	pongo pygma
8	11	100.0	110	Q6YK33	gorilla gor
9	11	100.0	110	AAP35454	homo sapi
10	11	100.0	110	AA06935	gorilla g
11	11	100.0	110	AA039451	homo sapi
12	7	63.6	110	INS_CANFA	canis fami
13	7	63.6	188	Q88AB3	pseudomora
14	7	63.6	456	O59732	schizosacch
15	7	63.6	497	ATRX_WACEU	macroopus eu
16	7	63.6	837	O7MTD5	porphyromon
17	7	63.6	1395	CG33_HUMAN	homo sapien
18	7	63.6	2592	Q84T10	oryza sativ
19	6	54.5	62	MLC2_HYLLA	hylobates l
20	6	54.5	62	MLC2_PANPA	pan paniscu
21	6	54.5	62	MLC2_PANTR	pan troglod
22	6	54.5	72	O9NRF0	homo sapien
23	6	54.5	78	MLC2_HUMAN	homo sapien
24	6	54.5	80	Q8Y9E7	listeria mo
25	6	54.5	80	Q722T5	listeria mo
26	6	54.5	80	AA034426	listeria
27	6	54.5	82	Q9BQC4	homo sapien
28	6	54.5	83	Q942Y3	oryza sativ
29	6	54.5	86	Q9BD10	homo sapien
30	6	54.5	86	Q9BQD1	homo sapien
31	6	54.5	86	Q89JU2	bradyrhizob

32 6 54.5 91 2 Q6Z8P0 Q6z8p0 oryza sativ
 33 6 54.5 91 2 BAD10060 Bad10060 oryza sat
 34 6 54.5 107 2 Q8DTN4 Q8dtn4 streptococ
 35 6 54.5 110 1 INS_CAVPO P01329 cavia porce
 36 6 54.5 110 1 INS_SPETR Q91x13 spermophilu
 37 6 54.5 117 1 GHR1_CANFA Q9bef8 canis fami
 38 6 54.5 117 2 BAC75929 Bac75929 canis fam
 39 6 54.5 121 2 Q99LS2 Q99ls2 mus musculu
 40 6 54.5 123 2 Q9KU85 Q9kj85 pseudomora
 41 6 54.5 135 2 Q7PEC2 Q7pec2 anopheles g
 42 6 54.5 146 2 Q6LB41 Q6lb41 oligotropha
 43 6 54.5 146 2 CAG28518 Cag28518 oligotrop
 44 6 54.5 149 2 Q72LH3 Q72lh3 thermus the
 45 6 54.5 149 2 AAS80434 Aas80434 thermus t

ALIGNMENTS

RESULT 1
 Q8HZ80 ID Q8HZ80 PRELIMINARY; PRT; 65 AA.
 AC Q8HZ80;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Insulin (Fragment)
 OS Pongo pygmaeus (Orangutan).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pongo.
 OX NCBI_TaxID=9600;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA O'Huigin C., Tichy H., Klein J.
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY092024; AAM76641.1;
 DR GO; GO:0005776; C:extracellular;
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological process; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 FT NON_TER 1
 FT NON_TER 65
 SQ SEQUENCE 65 AA; 6920 MW; B772017FD9BCABEA CRC64;

Query Match 100.0%; Score 11; DB 2; Length 65;
 Best Local Similarity 100.0%; Pred. No. 0.001;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLOK 11
 |||
 Db 48 QPLALEGSLOK 58

RESULT 2
 Q8HZ81 ID Q8HZ81 PRELIMINARY; PRT; 65 AA.
 AC Q8HZ81;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Insulin (fragment)
 OS Gorilla gorilla (Gorilla).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Gorilla.
 OX NCBI_TaxID=9593;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA O'Huigin C., Tichy H., Klein J.
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -!- SIMILARITY: Belongs to the insulin family.
 DR EMBL; AY092023; AAM76640.1; -.

DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological process; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR SMART; SM00078; ILGF; 1.
 KW Insulin family.
 FT NON_TER 1
 FT NON_TER 65
 FT SEQUENCE 65 AA; 6920 MW; B772017FD8BCABEA CRC64;
 SQ SEQUENCE 110 AA; 12019 MW; 95A1F54BE7B247F9 CRC64;

Query Match 100.0%; Score 11; DB 2; Length 65;
 Best Local Similarity 100.0%; Pred. No. 0.001;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
 |||||
 Db 48 QPLALEGSLQK 58

DR PROSITE; PS00262; INSULIN; 1.
 KW Direct protein sequencing; Glucose metabolism; Hormone;
 FT SIGNAL 1 24
 FT CHAIN 25 54 Insulin B chain.
 FT PROPEP 57 87 C peptide.
 FT CHAIN 90 110 Insulin A chain.
 FT DISULFID 31 96 Interchain.
 FT DISULFID 43 109 Interchain.
 FT DISULFID 95 100
 SQ SEQUENCE 110 AA; 12019 MW; 95A1F54BE7B247F9 CRC64;

Query Match 100.0%; Score 11; DB 1; Length 110;
 Best Local Similarity 100.0%; Pred. No. 0.0016;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
 |||||
 Db 78 QPLALEGSLQK 88

RESULT 3

INS_CERAE
 ID INS_CERAE STANDARD; PRT; 110 AA.
 AC P30407; P01309;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Insulin precursor.
 GN Name=INS;
 OS Cercopithecus aethiops (Green monkey) (Grivet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Cercopithecus.
 OX NCBI_TaxID=9534;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92219953; PubMed=1560757;
 RA Seino S., Bell G.I., Li W.;
 RA "Sequences of primate insulin genes support the hypothesis of a slower
 RT rate of molecular evolution in humans and apes than in monkeys.";
 RL Mol. Biol. Evol. 9:193-203(1992).
 RN [2]
 RP SEQUENCE OF 57-87.
 RX MEDLINE=72258016; PubMed=4626369;
 RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;
 RA "Determination of the amino acid sequence of the monkey, sheep, and
 RT dog proinsulin C-peptides by a semi-micro Edman degradation
 RP procedure";
 RL J. Biol. Chem. 247:4866-4871(1972).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X61092; CAA43405.1; -
 DR PIR; B42179; B42179.
 DR HSSP; P01308; 1A10.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; ILGF; 1.
 .PR

RESULT 4

INS_HUMAN
 ID INS_HUMAN STANDARD; PRT; 110 AA.
 AC P01308;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Insulin precursor.
 GN Name=INS;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80120725; PubMed=6243748;
 RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischler E.,
 RA Goodman H.M.;
 RA "Sequence of the human insulin gene.";
 RT Nature 284:26-32(1980).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80236313; PubMed=6248962;
 RA Ullrich A., Dull T.J., Gray A., Brosius J., Sures I.;
 RA "Genetic variation in the human insulin gene.";
 RL Science 209:612-615(1980).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80054779; PubMed=503234;
 RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,
 RA Rutter W.J.;
 RA "Nucleotide sequence of a cDNA clone encoding human proinsulin.";
 RL Nature 282:525-527(1979).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80147417; PubMed=6927840;
 RA Sures I., Goeddel D.V., Gray A., Ullrich A.;
 RA "Nucleotide sequence of human proinsulin complementary DNA.";
 RL Science 208:57-59(1980).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93364428; PubMed=8358440;
 RA Lucassen A.M., Bell J.I., Julier C., Lathrop M.;
 RA "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1
 RT kb segment of DNA spanning the insulin gene and associated VNTR.";
 RL Nat. Genet. 4:305-310(1993).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Pancreas;
 RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:22:24 ; Search time 11.6627 Seconds
(without alignments)
62.550 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLALEGSLOK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/1/aaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/aaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/aaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/aaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/aaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	53	100.0	23	US-08-105-904B-10	Sequence 10, Appl
2	53	100.0	23	US-08-114-877A-10	Sequence 10, Appl
3	53	100.0	23	US-08-537-928-10	Sequence 10, Appl
4	53	100.0	23	US-09-633-269-10	Sequence 10, Appl
5	53	100.0	27	US-08-508-664-17	Sequence 17, Appl
6	53	100.0	29	US-09-676-787-2	Sequence 2, Appl
7	53	100.0	32	5514646-5	Patent No. 5514646
8	53	100.0	33	US-09-639-667-24	Sequence 24, Appl
9	53	100.0	33	5514646-4	Patent No. 5514646
10	53	100.0	35	US-08-160-376A-3	Sequence 3, Appl
11	53	100.0	35	US-08-389-487-6	Sequence 6, Appl
12	53	100.0	35	US-09-134-836-3	Sequence 3, Appl
13	53	100.0	35	US-09-386-303A-3	Sequence 3, Appl
14	53	100.0	35	US-09-815-229-4	Sequence 4, Appl
15	53	100.0	35	US-09-947-563-3	Sequence 3, Appl
16	53	100.0	35	5457066-1	Patent No. 5457066
17	53	100.0	54	US-09-815-229-13	Sequence 13, Appl
18	53	100.0	86	US-07-715-183C-1	Sequence 1, Appl
19	53	100.0	86	US-09-477-924-2	Sequence 2, Appl
20	53	100.0	86	US-09-723-981-2	Sequence 2, Appl
21	53	100.0	86	US-09-723-896-2	Sequence 2, Appl
22	53	100.0	86	US-09-878-380-1	Sequence 1, Appl
23	53	100.0	90	US-08-030-731A-43	Sequence 43, Appl
24	53	100.0	91	US-09-676-787-7	Sequence 7, Appl
25	53	100.0	96	US-09-134-836-4	Sequence 4, Appl
26	53	100.0	96	US-09-134-836-5	Sequence 5, Appl
27	53	100.0	96	US-09-386-303A-4	Sequence 4, Appl

28	53	100.0	96	US-09-386-303A-5	Sequence 5, Appl
29	53	100.0	96	US-09-947-563-4	Sequence 4, Appl
30	53	100.0	96	US-09-947-563-5	Sequence 5, Appl
31	53	100.0	97	US-08-160-376A-4	Sequence 4, Appl
32	53	100.0	97	US-08-389-487-7	Sequence 7, Appl
33	53	100.0	97	US-09-099-307-6	Sequence 6, Appl
34	53	100.0	97	US-09-099-307-7	Sequence 7, Appl
35	53	100.0	97	US-09-099-307-8	Sequence 8, Appl
36	53	100.0	97	US-09-099-307-11	Sequence 11, Appl
37	53	100.0	98	US-09-701-968-7	Sequence 7, Appl
38	53	100.0	99	US-09-701-968-8	Sequence 8, Appl
39	53	100.0	100	US-09-701-968-9	Sequence 9, Appl
40	53	100.0	110	US-08-950-720A-11	Sequence 11, Appl
41	53	100.0	110	US-08-589-028-2	Sequence 2, Appl
42	53	100.0	110	US-08-589-028-4	Sequence 4, Appl
43	53	100.0	110	US-08-784-582-2	Sequence 2, Appl
44	53	100.0	110	US-08-784-582-4	Sequence 4, Appl
45	53	100.0	110	US-08-785-271-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
 US-08-105-904B-10
 ; Sequence 10, Application US/08105904B
 ; Patent No. 6001364
 ; GENERAL INFORMATION:
 ; APPLICANT: Rose, Keith
 ; APPLICANT: Offord, Robin
 ; TITLE OF INVENTION: HETERO-POLYOXIME COMPOUNDS AND THEIR
 ; PREPARATION BY PARALLEL ASSEMBLY
 ; TITLE OF INVENTION: HETERO-POLYOXIME COMPOUNDS AND THEIR
 ; PREPARATION BY PARALLEL ASSEMBLY
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
 ; STREET: 5 Palo Alto Square, 3000 El Camino Real
 ; CITY: Palo Alto
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 94306
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/105,904B
 ; FILING DATE: 31-AUG-1993
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/057,594
 ; FILING DATE: 05-MAY-1993
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Neeley, Richard L.
 ; REGISTRATION NUMBER: 30,092
 ; REFERENCE/DOCKET NUMBER: ABIC-001/02US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415)843-5000
 ; TELEFAX: (415)857-0663
 ; TELEX: 380816 COOLEYPA
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 23 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHEetical: NO
 ; FEATURE:
 ; NAME/KEY: Modified-site
 ; LOCATION: 1
 ; OTHER INFORMATION: AOA-Glu
 ; US-08-105-904B-10

Query Match 100.0%; Score 53; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.00037;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLQK 11
 |||||
 Db 12 QPLALEGSLQK 22

RESULT 2
 US-08-114-877A-10
 ; Sequence 10, Application US/08114877A
 ; Patent No. 6174530
 ; GENERAL INFORMATION:
 ; APPLICANT: Reese, Keith
 ; APPLICANT: Offord, Robin
 ; TITLE OF INVENTION: HOMOGENOUS POLYOXIME COMPOSITIONS AND THEIR
 ; TITLE OF INVENTION: PREPARATION BY PARALLEL ASSEMBLY
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
 ; STREET: 5 Palo Alto Square
 ; CITY: Palo Alto
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 94036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/114,877A
 ; FILING DATE: 31-AUG-1993
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/057,594
 ; FILING DATE: 05-MAY-1993
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Neeley, Richard L.
 ; REGISTRATION NUMBER: 30,092
 ; REFERENCE/DOCKET NUMBER: ABIC-001/01US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 843 5070
 ; TELEFAX: (415) 857-0663
 ; TELEX: 380816 CooleyPA
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 23 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHEITICAL: NO
 ; FEATURE:
 ; NAME/KEY: Modified-site
 ; LOCATION: 1
 ; OTHER INFORMATION: AOA-Glu
 ; US-08-114-877A-10

Query Match 100.0%; Score 53; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.00037;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLQK 11
 |||||
 Db 12 QPLALEGSLQK 22

RESULT 3
 US-08-537-928-10
 ; Sequence 10, Application US/08537928
 ; Patent No. 6217873
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; APPLICANT: Offord, Robin E
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GREN-001/03US
 ; CURRENT APPLICATION NUMBER: US/08/537,928
 ; CURRENT FILING DATE: 1996-01-05
 ; EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; EARLIER FILING DATE: 1994-05-05
 ; EARLIER APPLICATION NUMBER: 08/105,904
 ; EARLIER FILING DATE: 1993-08-31
 ; EARLIER APPLICATION NUMBER: 08/114,877
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 ; US-09-633-269-10

Query Match 100.0%; Score 53; DB 4; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.00037;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLQK 11
 |||||
 Db 12 QPLALEGSLQK 22

Query Match 100.0%; Score 53; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.00037;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLQK 11
 |||||
 Db 12 QPLALEGSLQK 22

RESULT 4
 US-09-633-269-10
 ; Sequence 10, Application US/09633269
 ; Patent No. 6663869
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; APPLICANT: Offord, Robin E
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GREN-001/03US
 ; CURRENT APPLICATION NUMBER: US/09/633,269
 ; CURRENT FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/08/537,928
 ; PRIOR FILING DATE: 1996-01-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1994-05-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/105,904
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1993-08-31
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/114,877
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 ; US-09-633-269-10

Query Match 100.0%; Score 53; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.00037;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLALEGSLQK 11
 |||||
 Db 12 QPLALEGSLQK 22

RESULT 5
 US-08-537-928-10
 ; Sequence 10, Application US/08537928
 ; Patent No. 6217873
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith, Rose
 ; APPLICANT: Offord, Robin E
 ; TITLE OF INVENTION: POLYOXIME COMPOSITIONS AND THEIR PREPARATION
 ; FILE REFERENCE: GREN-001/03US
 ; CURRENT APPLICATION NUMBER: US/08/537,928
 ; CURRENT FILING DATE: 1996-01-05
 ; EARLIER APPLICATION NUMBER: PCT/IB94/00093
 ; EARLIER FILING DATE: 1994-05-05
 ; EARLIER APPLICATION NUMBER: 08/105,904
 ; EARLIER FILING DATE: 1993-08-31
 ; EARLIER APPLICATION NUMBER: 08/114,877
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 23
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Aminoxyacetyl-Glu
 ; US-09-633-269-10

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:54:39 ; Search time 143.795 Seconds
(without alignments)
24,802 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLALEGSLOK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1370721 seqs, 324215800 residues

Total number of hits satisfying chosen parameters: 1370721

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Table with columns: Result No., Score, Query Match, Length, DB ID, Description. Contains 20 rows of sequence identifiers and their corresponding scores and lengths.

Table with columns: 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45. Contains sequence identifiers and their corresponding scores and lengths.

ALIGNMENTS

RESULT 1
US-10-617-568-23
; Sequence 23, Application US/10617568
; Publication No. US20040197862A1
; GENERAL INFORMATION:
; APPLICANT: Dana-Farber Cancer Institute, Inc. et al.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: GENERATION OF MAC CLASS II COMPOUNDS BY
; TITLE OF INVENTION: PEPTIDE EXCHANGE
; FILE REFERENCE: DFN-044
; CURRENT APPLICATION NUMBER: US/10/617,568
; CURRENT FILING DATE: 2003-07-11
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 60/395494
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 60/397893
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-617-568-23

Query Match 100.0%; Score 53; DB 17; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 QPLALEGSLOK 11
Db 6 QPLALEGSLOK 16
RESULT 2
US-10-641-834-24
; Sequence 24, Application US/10641834

```

; Publication No. US20040209830A1
; GENERAL INFORMATION: Stephen James
; APPLICANT: Russell, Stephen James
; APPLICANT: Peng, Kah Whye
; TITLE OF INVENTION: SYSTEM FOR MONITORING THE EXPRESSION OF
; TITLE OF INVENTION: TRANSGENES
; FILE REFERENCE: 07039-292002
; CURRENT APPLICATION NUMBER: US/10/641,834
; CURRENT FILING DATE: 2003-05-18
; PRIOR APPLICATION NUMBER: 09/639,667
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: 60/149,168
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: C-peptide sequence
US-10-641-834-24

```

```

Query Match 100.0%; Score 53; DB 17; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 QPLALEGSLQK 11
Db 22 QPLALEGSLQK 32

```

```

RESULT 3
US-09-815-229-4
; Sequence 4, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS DISORDERS
; FILE REFERENCE: P178681US
; CURRENT APPLICATION NUMBER: US/09/815,229
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 4
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-815-229-4

```

```

Query Match 100.0%; Score 53; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 QPLALEGSLQK 11
Db 24 QPLALEGSLQK 34

```

```

RESULT 4
US-09-947-563-3
; Sequence 3, Application US/09947563
; Patent No. US20020156234A1
; GENERAL INFORMATION:
; APPLICANT: Rubroder, Franz-Josef
; APPLICANT: Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; insulins precursors having correctly bonded cystine bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farrah, Garrett &

```

```

; Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/947,563
; FILING DATE: 07-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/134,836
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Leslie McDonnell
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02481.1600-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..35
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-947-563-3

```

```

Query Match 100.0%; Score 53; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 QPLALEGSLQK 11
Db 24 QPLALEGSLQK 34

```

```

RESULT 5
US-10-740-098-4
; Sequence 4, Application US/10740098
; Publication No. US20040138101A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS DISORDERS
; FILE REFERENCE: P178681US
; CURRENT APPLICATION NUMBER: US/10/740,098
; CURRENT FILING DATE: 2003-12-17
; PRIOR APPLICATION NUMBER: US/09/815,229
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 4
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-740-098-4

```

```

Query Match 100.0%; Score 53; DB 16; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0044;

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 11:41:36 ; Search time 721.759 Seconds
(without alignments)
720.721 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLALEGSLQK 11

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=cdn2_1/USPTO spool_p/US10783095/runat_26102004_185426_24805/app_query.fasta_1.796
-DB=GenEmbl -QFMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LCOEPL=0 -DOOPEXT=0
-UNITS=BITS -START=1 -END=1 -MATRIX=blcsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10783095 @CCN 1 1.6148 @runat_26102004_185426_24805 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSRBLCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Table with 4 columns: Result No., Score, Match Length, DB ID, Description. Contains 4 rows of search results.

Table with 5 columns: Hit No., Score, Match Length, DB ID, Description. Contains 45 rows of search results.

ALIGNMENTS

RESULT 1
AR077834
LOCUS AR077834 68 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 18 from patent US 5962267.
ACCESSION AR077834
VERSION AR077834.1 GI:10004580
KEYWORDS Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS Shin,H.-C., Chang,S.-G., Kim,D.-Y. and Kim,C.-S.
TITLE Proinsulin derivative and process for producing human insulin
JOURNAL Patent: US 5962267-A 18 05-OCT-1999;
FEATURES Location/Qualifiers
source 1..68
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Alignment Scores:
Pred. No.: 0.00628 Length: 68
Score: 53.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-783-095-9 (1-11) x AR077834 (1-68)
 QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
 DB 7 CAGCCGTTGGCGCTGGAGGTTCCCTGCAGAAA 39

RESULT 2
 AXI07018 87 bp DNA linear PAT 30-APR-2001
 LOCUS
 DEFINITION Sequence 1 from Patent WO0125278.
 ACCESSION AXI07018
 VERSION AXI07018.1 GI:13922551
 KEYWORDS synthetic construct
 ORGANISM synthetic construct
 artificial sequences.
 REFERENCE 1 Habermann,P., Ertl,J., Meiwes,J. and Seipke,G.
 AUTHORS C-peptide for the improved production of insulin and insulin
 TITLE analogues
 JOURNAL Patent: WO 0125278-A 1 12-APR-2001;
 FEATURES Aventis Pharma Deutschland GmbH (DE)
 source Location/Qualifiers
 1..87
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="DNA kodierend fur C-PEPTIDEVARIANTE"

ORIGIN

Alignment Scores:
 Pred. No.: 0.00821 Length: 87
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-783-095-9 (1-11) x AXI07018 (1-87)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
 DB 52 CAGCCGTTGGCGCTGGAGGTTCCCTGCAGAG 84

RESULT 3
 XXHSIN 105 bp DNA linear SYN 04-AUG-1995
 LOCUS
 DEFINITION Synthetic gene for human insulin C-chain.
 ACCESSION V01461 J02546
 VERSION V01461.1 GI:58303
 KEYWORDS artificial gene; gene synthesis.
 SOURCE synthetic construct
 ORGANISM artificial sequences.
 REFERENCE 1 (bases 1 to 105)
 AUTHORS Hsiung,H.M., Sung,W.L., Brousseau,R., Wu,R. and Narang,S.A.
 TITLE Synthesis of the human insulin gene, Part III. Chemical synthesis
 of 5'-phosphomonoester group containing deoxyribonucleotides
 by the modified phosphotriester method. Its application in the
 synthesis of seventeen fragments constituting human insulin C-chain
 DNA
 Nucleic Acids Res. 8 (23), 5753-5765 (1980)
 JOURNAL 81124297
 MEDLINE
 PUBMED 7008029
 FEATURES Location/Qualifiers
 1..105
 /organism="synthetic construct"
 /mol_type="other DNA"
 /db_xref="taxon:32630"
 <1..>105
 /note="(amino acids 31 to 65)"
 /codon_start=1
 /transl_table=11

CDS

US-10-783-095-9 (1-11) x AXI07018 (1-87)

/product="insulin C chain"
 /protein_id="CAA24707.1"
 /db_xref="GI:929915"
 /translation="RREAEDLQVGVVELGGGPGAGSLQPLALEGSLQKR"

ORIGIN

Alignment Scores:
 Pred. No.: 0.0101 Length: 105
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 12 Gaps: 0

US-10-783-095-9 (1-11) x XXHSIN (1-105)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
 DB 70 CAACCTTTGGCTTCGAGGGATCACTTCARAAAG 102

RESULT 4
 AXI179492 216 bp DNA linear PAT 31-AUG-2001
 LOCUS
 DEFINITION Sequence 87 from Patent WO0131037.
 ACCESSION AXI179492
 VERSION AXI179492.1 GI:14599120
 KEYWORDS synthetic construct
 SOURCE synthetic construct
 ORGANISM artificial sequences.

REFERENCE 1 Ben-Nun,A., Kerlero de Rosbo,N. and Sappier,G.P.
 AUTHORS Synthetic human genes and polypeptides and their use in the
 TITLE treatment of autoimmune diseases
 JOURNAL Patent: WO 0131037-A 87 03-MAY-2001;
 YEDA RESEARCH AND DEVELOPMENT CO. LTD. (IL)
 FEATURES Location/Qualifiers
 1..216
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 1..207
 /note="unnamed protein product; synthetic construct"

CDS

US-10-783-095-9 (1-11) x AXI179492 (1-216)

ORIGIN

Alignment Scores:
 Pred. No.: 0.0221 Length: 216
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-783-095-9 (1-11) x AXI179492 (1-216)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
 DB 163 CAACCGTTGGCGCTGGAGGGTCCCTGCAGAAAAG 195

RESULT 5

101821 252 bp ss-DNA linear PAT 21-MAY-1993
 LOCUS
 DEFINITION Sequence 6 from Patent US 4801536.
 ACCESSION 101821
 VERSION 101821.1 GI:269797
 KEYWORDS Unknown.

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 28, 2004, 11:40:21 ; Search time 143.398 seconds
(without alignments)
402.682 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLALEGLQK 11

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlp
-Q=/cgn2_1/USPTO_spoop_p/US10783095/runat_26102004_185425_24797/app_query.fasta_1.796
-DB=N Geneseg 23Sep04 -OPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=oct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pfco -NOEM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095@CGN_1_1241@runat_26102004_185425_24797 -NCPU=6 -ICPU=3
-NO_WMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseg 23Sep04:*
1: Genesegq1980s:*
2: Genesegq1990s:*
3: Genesegq2000s:*
4: Genesegq2001as:*
5: Genesegq2001bs:*
6: Genesegq2002as:*
7: Genesegq2002bs:*
8: Genesegq2003as:*
9: Genesegq2003bs:*
10: Genesegq2003cs:*
11: Genesegq2003ds:*
12: Genesegq2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Table with 5 columns: Result No., Score, Query Match, Length, DB ID, Description. Contains 6 rows of search results.

Table with 5 columns: ID, AAX86082 standard; DNA; 66 BP, 15-SEP-1999 (first entry), PCR primer used to produce chimeric genes of the invention. Contains 45 rows of sequence identifiers.

ALIGNMENTS

Table with 5 columns: ID, AAX86082 standard; DNA; 66 BP, 15-SEP-1999 (first entry), PCR primer used to produce chimeric genes of the invention. Contains 45 rows of alignment data.

PI Mueller JP, Matis LA, Wang Y;
 DR WPI; 1999-405113/34.
 XX
 XX Chimeric fusion proteins comprising immunodominant epitopes of glutamate
 PT decarboxylase and insulin, used to treat patients that have or are
 FT predicted to be at risk of developing Type I diabetes.
 XX
 PS Example; Page 62; 67pp; English.
 XX
 CC The specification describes chimeric fusion proteins comprising
 CC immunodominant epitopes of glutamate decarboxylase (GAD) and insulin. The
 CC chimeric fusion proteins comprises insulin chain B and individual peptide
 CC moieties consisting of at least one GAD65 peptide capable of eliciting a
 CC human T cell response, where the insulin chain B and GAD65 peptide are
 CC covalently linked and the chimeric fusion protein is capable of eliciting
 CC a human T cell response to insulin chain B and to each of the at least
 CC one GAD65 peptides. The chimeric fusion proteins are used to treat
 CC patients that are predicted to be at risk of developing Type I diabetes
 CC and those suffering from Type I diabetes. The fusion proteins can also be
 CC used to diagnose patients suffering from insulin dependent diabetes
 CC mellitus (IDDM) and/or Stiff Man syndrome. This is also useful for
 CC determining the suitability of patients as recipients of transplants of
 CC pancreatic cells or tissues. The present sequence represents a primer
 CC used to construct genes encoding the chimeric fusion proteins of the
 CC invention

XX
 SQ Sequence 66 BP; 10 A; 27 C; 21 G; 8 T; 0 U; 0 Other;

Alignment Scores: Length: 66
 Pred. No.: 0.0131 Matches: 11
 Score: 53.00
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-783-095-9 (1-11) x AMX86082 (1-66)
 QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
 |||||
 DB 42 CAGCCCTTGGCCCTGGAGGGTCCCTGCAGAAG 10

RESULT 2
 AAF61522
 ID AAF61522 standard; DNA; 87 BP.
 XX
 AC AAF61522;
 XX
 DT 04-JUL-2001 (first entry)
 XX
 DE Human proinsulin C-peptide encoding DNA.
 XX
 KW Insulin; proinsulin; C-peptide; human; B-chain; A-chain; diabetes; ds.
 XX
 OS Homo sapiens.
 XX
 FN DE19947456-A1.
 XX
 PD 05-APR-2001.
 XX
 PF 02-OCT-1999; 99DE-01047456.
 XX
 PR 02-OCT-1999; 99DE-01047456.
 XX
 PA (AVET) AVENTIS PHARMA DEUT GMBH.
 XX
 PI Habermann P, Ertl J, Meiwes J, Seipke G;
 XX
 DR WPI; 2001-267042/28.
 DR P-PSDB; AAB70847.
 XX
 PT New synthetic derivatives of the C-peptide of proinsulin, useful in the

PT preparation of human insulin or insulin analogs in high yield.
 XX
 PS Disclosure; Page 7; 12pp; German.

XX
 CC This invention describes novel synthetic derivatives (I) of the C-peptide
 CC of proinsulin. Such novel derivatives of human insulin or insulin analogs
 CC represented by (I) are of formula Fus-B(1-30)-RDVP-Yn-A(1-21). Fus = the
 CC optionally present fusion component of arbitrary sequence; B(1-30) = the
 CC B-chain of human insulin; Y = amino acid chain having a basic amino acid
 CC at the C-terminal; n = 2-50 (defining the chain length of Y), and A(1-21)
 CC = the A-chain of human insulin; the A- and/or B-chain(s) may be modified
 CC by amino acid exchanges, additions and/or deletions. (I) are
 CC intermediates for human insulin or its analogs, which are used in the
 CC treatment of diabetes. A claimed method of preparing human insulin (or
 CC analogs) involves preparing (I), folding (I) so that disulfide bonds as
 CC in human insulin can form, enzymatically removing the RDVP-Yn part and
 CC Fus (if present) and purifying the product. Insulin (or analogs) can be
 CC prepared in increased yields via (I). Specifically the expression yield
 CC can be increased around 20% compared with that obtained using plasmid
 CC pIN30d and almost 5-fold compared with that obtained using plasmid
 CC pIN302d. The control of enzymatic processing is also improved. The
 CC folding rate is comparable with that obtained using simian proinsulin
 CC encoded by pIN30d. This sequence encodes the human proinsulin C-peptide
 CC which is described in the method of the invention

XX
 SQ Sequence 87 BP; 10 A; 27 C; 37 G; 13 T; 0 U; 0 Other;

Alignment Scores: Length: 87
 Pred. No.: 0.0181 Matches: 11
 Score: 53.00
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-783-095-9 (1-11) x AAF61522 (1-87)
 QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
 |||||
 DB 52 CAGCCCTTGGCCCTGGAGGGTCCCTGCAGAAG 84

RESULT 3
 AAT45977
 ID AAT45977 standard; cDNA; 159 BP.
 XX
 AC AAT45977;
 XX
 DT 16-MAR-1997 (first entry)
 XX
 DE Oligonucleotide for 3' end of single chain insulin.
 XX
 KW Single chain insulin; diabetes mellitus; gene therapy; ss.
 XX
 OS Synthetic.
 XX
 FN EP741188-A2.
 XX
 PD 06-NOV-1996.
 XX
 PF 03-MAY-1996; 96EP-00303133.
 XX
 PR 05-MAY-1995; 95US-00435503.
 PR 05-MAY-1995; 95US-00435572.
 XX
 PA (ELIL) LILLY & CO ELI.
 XX
 PI Chance RE, Dimarchi RD, Hoffmann JA, Long HB, Miller AR;
 XX
 DR WPI; 1996-487391/49.
 XX
 PT Single chain insulin polypeptide(s) - used for treating diabetes.
 XX
 PS Example 7; Page 15; 22pp; English.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 11:54:47 ; Search time 32.6024 Seconds
(without alignments)
239.819 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLALEGSLQK 11

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 0%
Listing first 45 summaries

Post-processing: Minimum Match 0%

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10783095/runat_26102004_185427_24832/app_query.fasta_1.796
-DB-issued Patents NA -QFMT=fastap -SUFFIX=rrn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-NODE=LOCAL -OUTFMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=JUS10783095@cgn1_1_213@runat_26102004_185427_24832 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	53	100.0	68	2	US-08-600-783-18
2	53	100.0	87	4	US-09-676-787-1
3	53	100.0	258	2	US-08-600-783-17
4	53	100.0	269	2	US-08-673-312-10
5	53	100.0	275	1	US-07-918-953-12
6	53	100.0	275	1	US-07-918-953-14
7	53	100.0	275	1	US-08-081-661-12
8	53	100.0	275	1	US-08-081-661-14
9	53	100.0	276	1	US-07-918-953-16
10	53	100.0	276	1	US-08-081-661-16
11	53	100.0	277	1	US-07-715-183C-3
12	53	100.0	278	1	US-07-715-183C-2

13	53	100.0	281	1	US-07-764-655D-12
14	53	100.0	281	1	US-07-764-655D-13
15	53	100.0	281	6	5514646-1
16	53	100.0	290	1	US-08-030-731A-41
17	53	100.0	290	1	US-08-030-731A-42
18	53	100.0	298	1	US-07-826-928A-28
19	53	100.0	304	1	US-07-896-551B-12
20	53	100.0	330	4	US-08-472-701-1
21	53	100.0	330	5	PCT-US95-08596-1
22	53	100.0	351	4	US-09-280-030-49
23	53	100.0	359	3	US-08-589-028-3
24	53	100.0	359	3	US-08-784-582-3
25	53	100.0	359	3	US-08-785-271-3
26	53	100.0	390	4	US-09-280-030-48
27	53	100.0	416	3	US-08-945-140-5
28	53	100.0	450	4	US-09-185-852-1
29	53	100.0	450	4	US-09-323-738-1
30	53	100.0	510	1	US-07-918-953-7
31	53	100.0	510	1	US-08-081-661-7
32	53	100.0	515	3	US-08-589-028-1
33	53	100.0	515	3	US-08-784-582-1
34	53	100.0	515	3	US-08-785-271-1
35	53	100.0	2499	4	US-09-775-508C-7
36	53	100.0	4992	4	US-09-015-399-10
37	50	94.3	68	4	US-09-485-286-7
38	50	94.3	74	4	US-09-485-286-9
39	50	94.3	4646	4	US-09-485-286-14
40	48	90.6	51	6	5514646-42
41	48	90.6	51	6	5514646-43
42	47	88.7	598	3	US-09-537-696-11
43	47	88.7	633	3	US-09-537-696-12
44	47	88.7	666	3	US-09-537-696-13
45	39	73.6	99	1	US-07-918-953-3

ALIGNMENTS

RESULT 1
US-08-600-783-18 Application US/08600783
; Sequence 18
; GENERAL INFORMATION:
; APPLICANT: SHIN, Hang Cheol
; APPLICANT: CHANG, Seung Gu
; APPLICANT: KIM, Dae Young
; APPLICANT: KIM, Chong Suhli
; TITLE OF INVENTION: Proinsulin Derivative and Process
; TITLE OF INVENTION: for Producing Human Insulin
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SHIN, Hang Cheol
; STREET: Ssangma-Haunshin Apt. 102-1206,
; CITY: Kwangmyung-shi
; COUNTRY: Republic of Korea
; ZIP: 423-030
; ADDRESSEE: CHANG, Seung Gu
; STREET: Hyundai Apt. 71-203, Apkujung-dong,
; CITY: Kangnam-ku
; COUNTRY: Republic of Korea
; ZIP: 423-030
; ADDRESSEE: KIM, Dae Young
; STREET: Sosa Jukong Apt. 108-202, Sosa Bon-dong,
; CITY: Bucheon-shi
; COUNTRY: Republic of Korea
; ZIP: 422-230
; ADDRESSEE: KIM, Chong Suhli
; STREET: Garden Heights Apt. 202-801, #100,

```

; STREET: Hwangkeum-dong, Soosung-ku
; CITY: Taegu
; STATE: Taegu
; COUNTRY: Republic of Korea
; ZIP: 706-040
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk, 3.5 inch, 1.44MB storage
; COMPUTER: IBM PC/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,783
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: KR 95-2751
; FILING DATE: 15-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Shahan Islam
; REGISTRATION NUMBER: 32,507
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-1000
; TELEFAX: (212) 953-7249
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: oligonucleotide
; ANTI-SENSE: yes
US-08-600-783-18

```

```

Alignment Scores:
Pred. No.: 0.00108 Length: 68
Score: 53.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

```

```

US-10-783-095-9 (1-11) x US-08-600-783-18 (1-68)
Qy 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
Db 7 CAGCCGTTGGCGTGGAGGGTCCCTGCAGAAA 39

```

```

RESULT 2
US-09-676-787-1
; Sequence 1, Application US/09676787
; Patent No. 6534288
; GENERAL INFORMATION:
; APPLICANT: HABERMANN, Paul
; APPLICANT: ERTL, Johann
; APPLICANT: MEIWEIS, Johannes
; APPLICANT: SEIPKE, Gerhard
; TITLE OF INVENTION: C PEPTIDE FOR IMPROVED PREPARATION OF INSULIN AND INSULIN ANALOGS
; FILE REFERENCE: 38005-0110
; CURRENT APPLICATION NUMBER: US/09/676,787
; CURRENT FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: DE 199 47 456.7
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 87
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-676-787-1

```

```

Alignment Scores:
Pred. No.: 0.00146 Length: 87

```

```

Score: 53.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

```

```

US-10-783-095-9 (1-11) x US-09-676-787-1 (1-87)
Qy 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
Db 52 CAGCCGTTGGCGTGGAGGGTCCCTGCAGAG 84

```

```

RESULT 3
US-08-600-783-17
; Sequence 17, Application US/08600783
; Patent No. 5962267
; GENERAL INFORMATION:
; APPLICANT: SHIN, Hang Cheol
; APPLICANT: CHANG, Seung Gu
; APPLICANT: KIM, Dae Young
; APPLICANT: KIM, Chong Suhll
; TITLE OF INVENTION: Proinsulin Derivative and Process
; TITLE OF INVENTION: for Producing Human Insulin
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SHIN, Hang Cheol
; STREET: Ssangma-Hanshin Apt. 102-1206,
; STREET: #245 Cholsan-dong
; CITY: Kwangmyung-shi
; STATE: Kyungki-do
; COUNTRY: Republic of Korea
; ZIP: 423-030
; ADDRESSEE: CHANG, Seung Gu
; STREET: Hyundai Apt. 71-203, Apjukong-dong,
; STREET: Kangnam-ku
; CITY: Seoul
; STATE: Seoul
; COUNTRY: Republic of Korea
; ZIP: 135-110
; ADDRESSEE: KIM, Dae Young
; STREET: Sosa Jukong Apt. 108-202, Sosa Bon-dong,
; STREET: Sosa-ku
; CITY: Bucheon-shi
; STATE: Kyungki-do
; COUNTRY: Republic of Korea
; ZIP: 422-230
; ADDRESSEE: KIM, Chong Suhll
; STREET: Garden Heights Apt. 202-801, #100,
; STREET: Hwangkeum-dong, Soosung-ku
; CITY: Taegu
; STATE: Taegu
; COUNTRY: Republic of Korea
; ZIP: 706-040

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44MB storage
; COMPUTER: IBM PC/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,783
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: KR 95-2751
; FILING DATE: 15-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Shahan Islam
; REGISTRATION NUMBER: 32,507
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-1000
; TELEFAX: (212) 953-7249
; INFORMATION FOR SEQ ID NO: 17:

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 28, 2004, 13:47:02 ; Search time 127.494 Seconds
(without alignments)
442.402 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLALEGSLQK 11
Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3413475 segs, 2563800928 residues
Total number of hits satisfying chosen parameters: 6826950
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10783095/runat_26102004_185428_24911/app_query.fasta_1.796
-DB=Published Applications NA -OPMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=5 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NCGU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGJOB -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	53	100.0	258	15	US-10-383-285-1	Sequence 1, Appl
2	53	100.0	260	10	US-09-807-742-16	Sequence 16, Appl
3	53	100.0	260	10	US-09-807-742-17	Sequence 17, Appl
4	53	100.0	260	17	US-10-746-149-23	Sequence 23, Appl
5	53	100.0	260	18	US-10-609-019-21	Sequence 21, Appl
6	53	100.0	330	15	US-10-346-563-1	Sequence 1, Appl
7	53	100.0	330	16	US-10-321-717-1	Sequence 1, Appl
8	53	100.0	351	9	US-09-280-030-49	Sequence 49, Appl
9	53	100.0	390	9	US-09-804-409A-10	Sequence 48, Appl
10	53	100.0	450	9	US-09-969-748C-5	Sequence 10, Appl
11	53	100.0	450	10	US-10-328-813-1	Sequence 5, Appl
12	53	100.0	450	15	US-10-411-037-43	Sequence 1, Appl
13	53	100.0	450	16	US-10-411-037-43	Sequence 43, Appl
14	53	100.0	450	16	US-10-411-026-43	Sequence 43, Appl
15	53	100.0	450	16	US-10-410-962-43	Sequence 43, Appl
16	53	100.0	450	16	US-10-411-049-43	Sequence 43, Appl
17	53	100.0	450	17	US-10-410-930-43	Sequence 43, Appl
18	53	100.0	450	17	US-10-410-997-43	Sequence 43, Appl
19	53	100.0	450	17	US-10-411-012-43	Sequence 43, Appl
20	53	100.0	450	17	US-10-287-994-43	Sequence 43, Appl
21	53	100.0	450	17	US-10-410-913-43	Sequence 43, Appl
22	53	100.0	1182	9	US-10-419-344-27	Sequence 27, Appl
23	53	100.0	1626	18	US-10-609-019-27	Sequence 27, Appl
24	53	100.0	2499	10	US-09-775-508A-1	Sequence 1, Appl
25	53	100.0	10487	17	US-10-746-149-44	Sequence 44, Appl
26	53	100.0	10487	18	US-10-609-019-32	Sequence 32, Appl
27	53	100.0	10512	17	US-10-746-149-42	Sequence 42, Appl
28	53	100.0	10512	18	US-10-609-019-31	Sequence 31, Appl
29	53	100.0	10880	17	US-10-746-149-47	Sequence 47, Appl
30	53	100.0	10895	18	US-10-609-019-41	Sequence 41, Appl
31	53	100.0	11255	17	US-10-746-149-43	Sequence 43, Appl
32	53	100.0	11271	18	US-10-609-019-42	Sequence 42, Appl
33	53	100.0	11332	18	US-10-609-019-43	Sequence 43, Appl
34	50	94.3	68	16	US-10-430-752A-7	Sequence 7, Appl
35	50	94.3	74	16	US-10-430-752A-9	Sequence 9, Appl
36	50	94.3	4646	16	US-10-430-752A-14	Sequence 14, Appl
37	48	90.6	375	16	US-10-221-677-25	Sequence 25, Appl
38	48	90.6	1944	16	US-10-221-677-26	Sequence 26, Appl
39	47	88.7	333	17	US-10-419-539-2	Sequence 2, Appl
40	47	88.7	598	13	US-10-013-032-11	Sequence 11, Appl
41	47	88.7	633	13	US-10-013-032-12	Sequence 12, Appl
42	47	88.7	666	13	US-10-013-032-13	Sequence 13, Appl
43	47	88.7	1080	17	US-10-419-539-3	Sequence 3, Appl
44	47	88.7	1217	17	US-10-419-539-4	Sequence 4, Appl
45	45	84.9	1616	13	US-10-027-632-252243	Sequence 252243,

ALIGNMENTS

RESULT 1
US-10-383-285-1
; Sequence 1, Application US/10383285
; Publication No. US20030200566A1
; GENERAL INFORMATION:
; APPLICANT: STREATHFIELD, STEPHEN
; APPLICANT: HOWARD, JOHN
; TITLE OF INVENTION: PRODUCTION OF INSULIN AND INSULIN-LIKE PROTEINS IN PLANTS
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 10040
; CURRENT APPLICATION NUMBER: US/10/383,285
; CURRENT FILING DATE: 2003-05-21
; PRIOR APPLICATION NUMBER: 60/362,874
; PRIOR FILING DATE: 2002-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 258
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-383-285-1

Alignment Scores:
 Pred. No.: 0.0152 Length: 258
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 15 Gaps: 0

US-10-783-095-9 (1-11) x US-10-383-285-1 (1-258)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 |||||
 Db 160 CAGCCGCTGGCCCTCGAGGGTCCCTCCAGAAG 192

RESULT 2

US-09-807-742-16
 ; Sequence 16, Application US/09807742
 ; Publication No. US20030204864A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DANIELL, HENRY
 ; TITLE OF INVENTION: PRODUCTION OF PHARMACEUTICAL PROTEINS IN TRANSGENIC
 ; TITLE OF INVENTION: PLASIDS
 ; FILE REFERENCE: 1465-PCT-US-00
 ; CURRENT APPLICATION NUMBER: US/09/807,742
 ; CURRENT FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: PCT/US01/06288
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 16
 ; LENGTH: 260
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-807-742-16

Alignment Scores:
 Pred. No.: 0.0153 Length: 260
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-783-095-9 (1-11) x US-09-807-742-16 (1-260)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 |||||
 Db 160 CAGCCGCTGGCCCTCGAGGGTCCCTCCAGAAG 192

RESULT 3

US-09-807-742-17
 ; Sequence 17, Application US/09807742
 ; Publication No. US20030204864A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DANIELL, HENRY
 ; TITLE OF INVENTION: PRODUCTION OF PHARMACEUTICAL PROTEINS IN TRANSGENIC
 ; TITLE OF INVENTION: PLASIDS
 ; FILE REFERENCE: 1465-PCT-US-00
 ; CURRENT APPLICATION NUMBER: US/09/807,742
 ; CURRENT FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: PCT/US01/06288
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 17
 ; LENGTH: 260
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Chloroplast
 ; OTHER INFORMATION: modified proinsulin sequence
 ; US-09-807-742-17

Alignment Scores:
 Pred. No.: 0.0153 Length: 260
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-783-095-9 (1-11) x US-09-807-742-17 (1-260)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 |||||
 Db 160 CAGCCGCTGGCCCTCGAGGGTCCCTCCAGAAG 192

RESULT 4

US-10-746-149-23
 ; Sequence 23, Application US/10746149
 ; Publication No. US20040172667A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cooper, Richard K.
 ; APPLICANT: Fioretti, William C.
 ; APPLICANT: Cadd, Gary G.
 ; TITLE OF INVENTION: Administration of Transposon-Based Vectors to Reproductive Organs
 ; FILE REFERENCE: 51667-0280 (51667-294923)
 ; CURRENT APPLICATION NUMBER: US/10/746,149
 ; CURRENT FILING DATE: 2003-12-24
 ; PRIOR APPLICATION NUMBER: US 60/392,415
 ; PRIOR FILING DATE: 2002-06-26
 ; PRIOR APPLICATION NUMBER: US 60/441,381
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,447
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,405
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,502
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,392
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 60/441,377
 ; PRIOR FILING DATE: 2003-01-21
 ; PRIOR APPLICATION NUMBER: US 10/609,019
 ; PRIOR FILING DATE: 2003-06-26
 ; NUMBER OF SEQ ID NOS: 52
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 23
 ; LENGTH: 260
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Synthetic
 ; US-10-746-149-23

Alignment Scores:
 Pred. No.: 0.0153 Length: 260
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 17 Gaps: 0

US-10-783-095-9 (1-11) x US-10-746-149-23 (1-260)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnLys 11
 |||||
 Db 160 CAGCCGCTGGCCCTCGAGGGTCCCTCCAGAAG 192

RESULT 5

US-10-609-019-21
 ; Sequence 21, Application US/10609019
 ; Publication No. US20040197910A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cooper, Richard K.
 ; US-10-609-019-21

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: October 28, 2004, 10:15:44 ; Search time 8.87952 Seconds
(without alignments)
119.194 Million cell updates/sec

Title: US-10-783-095-9
Perfect score: 53
Sequence: 1 QPLALEGSLQK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	53	100.0	110	1	IPHU
2	53	100.0	110	2	insulin precursor
3	53	100.0	110	2	insulin precursor
4	53	100.0	110	2	insulin precursor
5	50	94.3	110	1	IPDG
6	41	77.4	110	1	IPGP
7	41	77.4	128	2	Ycf35 protein - Sy
8	36	67.9	156	2	Glutamine-fructose
9	36	67.9	210	2	hypothetical prote
10	36	67.9	374	2	integrase - Lactoc
11	36	67.9	374	2	prophage pi2 prote
12	36	67.9	374	2	integrase - phage
13	36	67.9	597	2	Glutamine-fructose
14	36	67.9	597	2	Glutamine-fructose
15	36	67.9	598	2	Glutamine-fructose
16	36	67.9	608	2	Glutamine-fructose
17	36	67.9	608	2	hypothetical prote
18	36	67.9	609	2	protease IV [imp
19	35	66.0	1084	2	type II restriclio
20	35	66.0	1084	2	type II site-speci
21	34	64.2	110	1	INRB
22	34	64.2	388	2	Hsp70 cofactor [im
23	34	64.2	456	2	hypothetical prote
24	34	64.2	541	2	Glutamine-fructose
25	34	64.2	584	2	hypothetical prote
26	34	64.2	591	2	hypothetical prote
27	34	64.2	592	2	Glutamine-fructose
28	34	64.2	605	2	Glutamine-fructose
29	34	64.2	608	1	S01040

30	64.2	608	2	B95322
31	64.2	609	1	XNECGM
32	64.2	609	2	AB0955
33	64.2	609	2	B81528
34	64.2	609	2	F86611
35	64.2	609	2	E72012
36	64.2	609	2	A82844
37	64.2	609	2	H86058
38	64.2	609	2	G91212
39	64.2	611	2	T45493
40	64.2	612	2	B81246
41	64.2	612	2	H82022
42	64.2	612	2	AD3595
43	64.2	631	2	S74575
44	64.2	1896	2	T08851
45	64.2	1978	2	S77257

ALIGNMENTS

RESULT 1

IPHU
insulin precursor [validated] - human
N;Alternate names: preproinsulin
C;Species: Homo sapiens (man)
C;Date: 23-Oct-1981 #sequence, revision 23-Oct-1981 #text change 09-Jul-2004
C;Accession: A93222; A94253; A93216; A94251; A93144; A92075; A91186; I58114; A01579; S58
R;Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Cordell, B.; Tischer, E.; Goodman, H.M.
Nature 284, 26-32, 1980
A;Title: Sequence of the human insulin gene.
A;Reference number: A93222; MUID:80120725; PMID:6243748
A;Accession: A93222
A;Molecule type: DNA
A;Residues: 1-110 <BEL>
A;Cross-references: UNIPROT:P01308; GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
R;Ullrich, A.; Dall, T.J.; Gray, A.; Brosius, J.; Sures, J.
Science 209, 612-615, 1980
A;Title: Genetic variation in the human insulin gene.
A;Reference number: A94253; MUID:80236313; PMID:6248962
A;Accession: A94253
A;Molecule type: DNA
A;Residues: 1-110 <ULL>
A;Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
R;Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.
Nature 282, 525-527, 1979
A;Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.
A;Reference number: A93216; MUID:80054779; PMID:503234
A;Accession: A93216
A;Molecule type: mRNA
A;Residues: 1-110 <BEL2>
A;Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
R;Sures, J.; Goeddel, D.V.; Gray, A.; Ullrich, A.
Science 208, 57-59, 1980
A;Title: Nucleotide sequence of human preproinsulin complementary DNA.
A;Reference number: A94251; MUID:80147417; PMID:6927840
A;Accession: A94251
A;Molecule type: mRNA
A;Residues: 1-110 <SUR>
A;Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
R;Nicol, D.S.H.W.; Smith, L.F.
Nature 187, 483-485, 1960
A;Title: Amino-acid sequence of human insulin.
A;Reference number: A93144
A;Accession: A93144
A;Molecule type: protein
A;Residues: 25-54;90-110 <NIC>
R;Over, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.
J. Biol. Chem. 246, 1375-1386, 1971
A;Title: Studies on human preproinsulin. Isolation and amino acid sequence of the human pan
A;Reference number: A92075; MUID:7116410; PMID:5101771
A;Accession: A92075
A;Molecule type: protein

A;Residues: 57-87 <OYE>
 R;Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
 Eur. J. Biochem. 20, 190-199, 1971
 A;Title: Amino acid sequence of the C-peptide of human proinsulin.
 A;Reference number: A91186; MUID:71257722; PMID:5560404
 A;Accession: A91186
 A;Molecule type: protein
 A;Residues: 57-87 <SKOA>
 R;Lucassen, A.M.; Jullier, C.; Beressif, J.P.; Boitard, C.; Froguel, P.; Lathrop, M.; Bell
 Nature Genet. 4, 305-310, 1993
 A;Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment
 A;Reference number: I58114; MUID:93364428; PMID:8358440
 A;Accession: I58114
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-59,63-110 <RES>
 A;Cross-references: GB:L15440; NID:9307071; PIDN:AAA59179.1; PID:Q307072
 R;Sieber, P.; Kamber, B.; Hartmann, A.; Joehl, A.; Riniker, B.; Rittel, W.
 Helv. Chim. Acta 57, 2617-2621, 1974
 A;Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindungen.
 A;Reference number: A91636; MUID:75077277; PMID:4443293
 A;Contents: annotation; synthesis
 A;Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was identical
 A;Note: article in German with English abstract
 R;Naithani, V.K.
 Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973
 A;Title: The synthesis of C-peptide of human proinsulin.
 A;Reference number: A91658; MUID:75040007; PMID:4803504
 A;Contents: annotation; synthesis of residues 57-87
 R;Geiger, R.; Jaeger, G.; Koenig, W.
 Chem. Ber. 106, 2347-2352, 1973
 A;Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9
 A;Reference number: A90914
 A;Contents: annotation; synthesis of residues 57-87
 R;Kaufmann, J.E.; Irmingier, J.C.; Halban, P.A.
 Biochem. J. 310, 869-874, 1995
 A;Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide juncti
 A;Reference number: S58661; MUID:96013185; PMID:7575420
 A;Contents: annotation; site-directed mutagenesis study of proteolytic processing
 C;Genetics:
 A;Gene: GDB:INS
 A;Cross-references: GDB:119349; OMIM:176730
 A;Map position: 11p15.5-11p15.5
 A;Introns: 63/1
 C;Superfamily: insulin
 C;Keywords: hormone; pancreas
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-54/Domain: insulin chain B #status experimental <BCH>
 F;57-87/Domain: insulin #status experimental <MAT>
 F;90-110/Domain: connecting C peptide #status experimental <CPEP>
 F;91-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 100.0%; Score 53; DB 1; Length 110;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPAALEGSLOK 11
 |||||
 Db 78 QPAALEGSLOK 88

RESULT 2
 B42179
 insulin precursor - green monkey
 C;Species: Cercopithecus aethiops (green monkey, grivet)
 C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C;Accession: B42179; A05232; S16494; S22056
 R;Saino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of m
 A;Reference number: A42179; MUID:92219953; PMID:1560757
 A;Accession: B42179

A;Molecule type: DNA
 A;Residues: 1-110 <SEI>
 A;Cross-references: UNIPROT:P30407; EMBL:X61092; NID:922808; PIDN:CAA43405.1; PID:922809
 A;Note: sequence extracted from NCBI backbone (NCBIN:95185, NCBIP:95194)
 R;Peterson, J.D.; Nehrllich, S.; Oyer, P.E.; Steiner, D.F.
 J. Biol. Chem. 247, 4866-4871, 1972
 A;Title: Determination of the amino acid sequence of the monkey, sheep, and dog proinsulin;
 A;Reference number: A92111; MUID:72258016; PMID:4626369
 A;Molecule type: protein
 A;Accession: A05232
 A;Residues: 57-87 <PEI>
 C;Genetics:
 A;Introns: 63/1
 C;Superfamily: insulin
 C;Keywords: hormone; pancreas
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-54/Domain: insulin chain B #status predicted <BCH>
 F;57-87/Domain: insulin #status predicted <MAT>
 F;90-110/Domain: connecting peptide #status experimental <CPEP>
 F;91-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 100.0%; Score 53; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPAALEGSLOK 11
 |||||
 Db 78 QPAALEGSLOK 88

RESULT 3
 JQ0178
 insulin precursor - crab-eating macaque
 C;Species: Macaca fascicularis (crab-eating macaque)
 C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
 C;Accession: JQ0178
 R;Wettkam, W.; Gronenberg, J.; Leineweber, M.; Wengenmayer, F.; Winnacker, E.L.
 Gene 19, 179-183, 1982
 A;Title: The nucleotide sequence of cDNA coding for proinsulin from the primate Macaca
 A;Reference number: JQ0178; MUID:83080474; PMID:6184262
 A;Accession: JQ0178
 A;Molecule type: mRNA
 A;Residues: 1-110 <WET>
 A;Cross-references: UNIPROT:P30406; GB:J00336; NID:9342121; PIDN:AAA36849.1; PID:9342122
 C;Superfamily: insulin
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-54,90-110/Product: insulin #status predicted <MAT>
 F;57-87/Domain: insulin chain B #status predicted <BCH>
 F;90-110/Domain: insulin connecting C peptide #status predicted <CPT>
 F;91-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 100.0%; Score 53; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPAALEGSLOK 11
 |||||
 Db 78 QPAALEGSLOK 88

RESULT 4
 A42179
 insulin precursor - chimpanzee
 C;Species: Pan troglodytes (chimpanzee)
 C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C;Accession: A42179; S22058
 R;Saino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of m
 A;Reference number: A42179; MUID:92219953; PMID:1560757
 A;Accession: A42179

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 11:54:21 ; Search time 1251.08 Seconds
(without alignments)
320.392 Million cell updates/sec

Title: US-10-783-095-9

Perfect score: 53

Sequence: 1 QPLALEGSLQK 11

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlp

-MODEL=frame_p2n.model -DEV=xlp
-Q/cgn2_l/USPTO_spoor_p/US10783095/runat_26102004_185426_24817/app_query.fasta_1.796
-DB=EST -QFMT=fastp -SUFFIX=first -MINMATCH=0.1 -LCOEFL=0 -LCOEEXT=0
-UNITG=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -NODE=LOCAL
-OUTFMT=ptc -NOSM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095@cgn 1 1 9525 @runat 26102004 185426 24817 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*

- 1: gb_est1:*
2: gb_est2:*
3: gb_hcc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Table with columns: Result No., Score, Query Match, Length, DB ID, Description. Contains 9 rows of search results.

Table with columns: C, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45. Contains alignment data for various sequences.

ALIGNMENTS

RESULT 1
BQ267254/c
LOCUS BQ267254.1
DEFINITION BQ267254.1 118 bp mRNA linear EST 15-JUL-2003
ik01b07.y1 Human insulinoma.Homo sapiens cDNA clone IMAGE:5779548
S, similar to SW:INS_HUMAN F01308 INSULIN PRECURSOR. [I], mRNA
sequence.

ACCESSION BQ267254
VERSION BQ267254.1
KEYWORDS GI:204923119

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 118)

AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,J., Marra,M., Page,D., Wylie,T., Martin,J., Blustain,A., Schmitt,A., Theising,B., Ritter,B., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,K., Tsagarisvillii,R., Williams,T., Jackson,Y. and Bowers,Y.

Endocrine Pancreas Consortium

Unpublished (2000)

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@hph.harvard.edu

Library was constructed by Dr. J. Ferrer. In vivo mass-excised to

pLibrary SK- by Dr. H. Inoue DNA sequencing by: Washington

University genome Sequencing Center For information on obtaining a

clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
 Possible reversed clone: similarity on wrong strand
 Seq primer: -40RP from Gibco.

FEATURES

source

```

1. .118
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5779548"
/tissue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Human insulinoma"
/note="Organ: pancreas; Vector: pBluescript SK-; Site 1: XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to pBluescript SK- by Dr. H. Inoue following the Washington University protocol (http://genome.wustl.edu/est/lambda_protocol.shtml). Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."

```

ORIGIN

Alignment Scores:
 Pred. No.: 0.324 Length: 118
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservatave: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 5 Gaps: 0

US-10-783-095-9 (1-11) x BQ267254 (1-118)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
 |||
 DB 113 CAGCCCTTGGCCCTGGAGGGTCCCTCCAGAG 81

RESULT 2
 BQ272508 118 bp mRNA EST 15-JUL-2003
 LOCUS ik01b07.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5779548
 DEFINITION 3', similar to SW:INS_HUMAN P01308 INSULIN PRECURSOR. [1] ; mRNA sequence.

ACCESSION BQ272508.1 GI:20497577
 VERSION BQ272508.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 118)
 AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scaerac,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarisvili,R., Williams,T., Jackson,Y. and Bowers,Y.

TITLE Endocrine Pancreas Consortium
 JOURNAL Unpublished (2000)
 COMMENT Other ESTs: ik01b07.y1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. J. Ferrer in vivo mass-excised to pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a

clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
 Possible reversed clone: similarity on wrong strand
 Seq primer: -40UP from Gibco.

FEATURES

source

```

1. .118
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5779548"
/tissue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Human insulinoma"
/note="Organ: pancreas; Vector: pBluescript SK-; Site 1: XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to pBluescript SK- by Dr. H. Inoue following the Washington University protocol (http://genome.wustl.edu/est/lambda_protocol.shtml). Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."

```

ORIGIN

Alignment Scores:
 Pred. No.: 0.324 Length: 118
 Score: 53.00 Matches: 11
 Percent Similarity: 100.00% Conservatave: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 5 Gaps: 0

US-10-783-095-9 (1-11) x BQ272508 (1-118)

QY 1 GlnProLeuAlaLeuGluGlySerLeuGlnIlys 11
 |||
 DB 5 CAGCCCTTGGCCCTGGAGGGTCCCTCCAGAG 37

RESULT 3
 BU073408/118 bp mRNA linear EST 27-AUG-2002
 LOCUS im35e06.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:6036971
 DEFINITION 3', similar to SW:INS_HUMAN P01308 INSULIN PRECURSOR. [1] ; mRNA sequence.

ACCESSION BU073408
 VERSION BU073408.1 GI:22514597
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 118)
 AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scaerac,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarisvili,R., Williams,T., Jackson,Y. and Bowers,Y.

TITLE Endocrine Pancreas Consortium
 JOURNAL Unpublished (2000)
 COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. J. Ferrer in vivo mass-excised to pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a
 clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)

Q38608 lactococcus
 Q9b016 bacterioph
 Q9cgt4 lactococcus
 Q7mau3 wolineella s
 Q2j94 h glucosami
 Q26060 h glucosami
 Q9pmt4 c glucosami
 Q7x5j2 helicobacte
 Q8ueh1 a glucosami
 P59362 b glucosami
 P94323 bradyrhizob
 Q6sey6 uncultured
 Q72sf2 desulfovibr
 Aar38436 unculture

32 36 67.9 374 2 Q38608
 33 36 67.9 374 2 Q9b016
 34 36 67.9 374 2 Q9cgt4
 35 36 67.9 595 2 Q7mau3
 36 36 67.9 596 1 GLMS_HELPJ
 37 36 67.9 596 1 GLMS_HELPY
 38 36 67.9 597 1 GLMS_CANOE
 39 36 67.9 597 2 Q7X5J2
 40 36 67.9 607 1 GLMS_AGR5
 41 36 67.9 607 1 GLMS_BRAJA
 42 36 67.9 607 1 NODM_BRAJA
 43 36 67.9 607 2 Q6SEV6
 44 36 67.9 607 2 Q72sf2
 45 36 67.9 607 2 AAR38436

ALIGNMENTS

RESULT 1
 Q8HZ80 PRELIMINARY; PRT; 65 AA.
 ID AC Q8HZ80
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Insulin (fragment).
 OS Pongo pygmaeus (Orangutan).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pongo.
 OX NCBI_TaxID=9600;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA O'Huigin C., Tichy H., Klein J.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY092024; AAM76641.1;
 DR GO; GO:0005179; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological process; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 FT NON TER 1
 FT NON TER 65
 SQ SEQUENCE 65 AA; 6920 MW; B772017FD8BCABEA CRC64;

Query Match Similarity 100.0%; Score 53; DB 2; Length 65;
 Best Local Similarity 100.0%; Pred. No. 0.006;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 QPLALEGSLOK 11
 |||||
 Db 48 QPLALEGSLOK 58

RESULT 2
 Q8HZ81 PRELIMINARY; PRT; 65 AA.
 ID AC Q8HZ81
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Insulin (fragment).
 OS Gorilla gorilla (gorilla).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Gorilla.
 OX NCBI_TaxID=9593;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA O'Huigin C., Tichy H., Klein J.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -|- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -|- SIMILARITY: Belongs to the insulin family.
 DR EMBL; AY020203; AAM76640.1; -.

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
 Run on: October 28, 2004, 06:48:48 ; Search time 38.9639 Seconds
 (without alignments)
 162.436 Million cell updates/sec

Title: US-10-783-095-9
 Perfect score: 53
 Sequence: 1 QPLALEGSLOK 11

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues 1825181
 Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : UniProt_02.*
 1: uniprot_sprot.*
 2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	65	Q8HZ80	pongo pygma
2	53	100.0	65	Q8HZ81	gorilla gor
3	53	100.0	110	INS_CERAE	cercopithe
4	53	100.0	110	INS_HUMAN	pongo pygma
5	53	100.0	110	INS_MACFA	macaca fasc
6	53	100.0	110	INS_PANTR	pan troglod
7	53	100.0	110	INS_PONPY	pongo pygma
8	53	100.0	110	Q6YK33	gorilla gor
9	53	100.0	110	AAP35454	homo sapi
10	53	100.0	110	AAR06935	gorilla g
11	53	100.0	110	AAM39451	homo sapi
12	50	94.3	110	INS_CANFA	canis fami
13	41	77.4	110	INS_CAVPO	cavia porce
14	41	77.4	110	INS_SPSTR	spermophi
15	41	77.4	128	Q55981	synchocyst
16	40	75.5	1164	Q72CD0	desulfovibr
17	40	75.5	1164	AAS95831	desulfovibr
18	39	73.6	188	Q88AB3	pseudomonas
19	38	71.7	670	Q96E24	homo sapien
20	38	71.7	675	Q8BI66	mus musculu
21	38	71.7	679	Q8TF50	homo sapien
22	37	69.8	117	GHRL_CANFA	canis fami
23	37	69.8	117	BAC75929	canis fam
24	36	67.9	110	Q9BYA1	homo sapien
25	36	67.9	110	GLMWNW6	felis silve
26	36	67.9	156	GLMS_SPHYA	glucosami
27	36	67.9	210	P91262	caenorhabd
28	36	67.9	358	Q17855	prochloroco
29	36	67.9	374	Q38088	bacterioph
30	36	67.9	374	Q38313	bacterioph
31	36	67.9	374	Q38325	lactococcus

DR GO: GO:0005576; C:extracellular; IEA.
 DR GO: GO:0005179; F:hormone activity; IEA.
 DR GO: GO:0007582; P:physiological process; IEA.
 DR InterPro: IPR004825; Ins/IGF/relax.
 DR Pfam: PF00049; Insulin; 1.
 DR SMART: SM00078; IIGF; 1.
 KW Insulin family.
 FT NON_TER 1
 FT NON_TER 65
 FT NON_TER 85
 SQ SEQUENCE 65 AA; 6920 MW; B772017FD8BCABEA CRC64;
 Query Match 100.0%; Score 53; DB 2; Length 65;
 Best Local Similarity 100.0%; Pred. No. 0.006;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
 |||||
 Db 48 QPLALEGSLQK 58

RESULT 3
 INS_CERAE STANDARD; PRT; 110 AA.
 AC P30407; P01309;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Insulin precursor.
 GN Name=INS;
 OS Cercopithecus aethiops (Green monkey) (Grivet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Cercopithecidae;
 OX NCBI_TaxID=9534;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92219953; PubMed=1560757;
 RA Saino S., Bell G.I., Li W.;
 RT "Sequences of primate insulin genes support the hypothesis of a slower
 RT rate of molecular evolution in humans and apes than in monkeys.";
 RL Mol. Biol. Evol. 9:193-203(1992).
 RN [2]
 RP SEQUENCE OF 57-87.
 EX MEDLINE=72258016; PubMed=4626369;
 RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;
 RT "Determination of the amino acid sequence of the monkey, sheep, and
 RT dog proinsulin C-peptides by a semi-micro Edman degradation
 RT procedure.";
 RL J. Biol. Chem. 247:4866-4871(1972).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X61092; CAA43405.1;
 CC PIR; B42179; B42179.
 CC HSP; P01308; IAI0.
 CC InterPro: IPR004825; Ins/IGF/relax.
 CC Pfam: PF00049; Insulin; 1.
 CC PRINTS; PR00277; INSULIN.
 CC SMART; SM00078; IIGF; 1.
 CC

DR PROSITE; PS00262; INSULIN; 1.
 KW Direct protein sequencing; Glucose metabolism; Hormone;
 KW Insulin family; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 Insulin B chain.
 FT PROPEP 57 87 C-peptide.
 FT CHAIN 90 110 Insulin A chain.
 FT DISULFID 31 96 Interchain.
 FT DISULFID 43 109 Interchain.
 FT DISULFID 95 100 Interchain.
 SQ SEQUENCE 110 AA; 12019 MW; 95A1F54BE7B247F9 CRC64;
 Query Match 100.0%; Score 53; DB 1; Length 110;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLALEGSLQK 11
 |||||
 Db 78 QPLALEGSLQK 88

RESULT 4
 INS_HUMAN STANDARD; PRT; 110 AA.
 AC P01308;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Insulin precursor.
 GN Name=INS;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80120725; PubMed=6243748;
 RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischler E.,
 RA Goodman H.M.;
 RT "Sequence of the human insulin gene.";
 RL Nature 284:26-32(1980).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80236313; PubMed=6248962;
 RA Ullrich A., Dull T.J., Gray A., Brosius J., Sures I.;
 RT "Genetic variation in the human insulin gene.";
 RL Science 209:612-615(1980).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80054779; PubMed=503234;
 RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,
 RA Rutter W.J.;
 RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin.";
 RL Nature 282:525-527(1979).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80147417; PubMed=6927840;
 RA Sures I., Goedel D.V., Gray A., Ullrich A.;
 RT "Nucleotide sequence of human preproinsulin complementary DNA.";
 RL Science 208:57-59(1980).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93364428; PubMed=8358440;
 RA Lucassen A.M., Bell J.J., Jullier C., Lathrop M.;
 RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1
 RT kb segment of DNA spanning the insulin gene and associated VNTR.";
 RL Nat. Genet. 4:305-310(1993).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:55:55 ; Search time 66.506 Seconds
(without alignments)
129.454 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 LKRVESFSPRSDYINASPIEHDP 24

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : A_Geneseq_23sep04:*
1: Genesep1980s:*
2: Genesep1990s:*
3: Genesep2000s:*
4: Genesep2001s:*
5: Genesep2002s:*
6: Genesep2003as:*
7: Genesep2003bs:*
8: Genesep2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Table with columns: Result No., Score, % Match, Query Length, DB ID, Description. Lists search results for various human and mouse proteins like Aay33360, Aay49854, etc.

Table with columns: ID, AAY33360 standard; protein; 294 AA., 24, 100.0, 979, 4, AAG80205, AAG80205 Human aut, etc.

ALIGNMENTS

RESULT 1
AAY33360
ID AAY33360 standard; protein; 294 AA.
XX
AC AAY33360;
XX
AC AAY33360; (first entry)
DT 29-NOV-1999 (first entry)
DE Human islet cell antigen clone ICA-512 protein fragment.

XX Islet cell antigen; ICA; human; pancreas; antibody; IDDM; immunoassay;
XX insulin-dependent type I diabetes mellitus; diagnosis; immunoglobulin;
XX T-cell; B-cell; islet cell destruction; immunoreactivity; disease;
XX immune therapy; stimulating antigen; screening.

OS Homo sapiens.
XX
XX US955345-A.
XX
XX 21-SEP-1999.
XX
XX 06-JUN-1995; 95US-00468576.
XX
XX 17-FEB-1989; 89US-00312543.
XX
XX 04-DEC-1989; 89US-00441703.
XX
XX 14-JUN-1991; 91US-00715181.
XX
XX 08-JUN-1992; 92US-00872646.
XX
XX 05-MAY-1994; 94US-00239276.
XX
XX (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
XX
XX Rabin DU;
XX
XX WPI; 1999-560494/47.
XX
XX N-PSDB; AAZ09922.

XX Isolated DNA encoding antigens reactive with serum from diabetics useful
XX to develop products for the diagnosis and monitoring of insulin-dependent
XX diabetes mellitus.
XX
XX Example; Col 85-87; 63pp; English.
XX
XX This invention describes novel isolated and purified DNAs encoding
XX pancreatic islet cell antigens reactive with islet cell antibodies found
XX in sera of patients with insulin-dependent type I diabetes mellitus
XX (IDDM). The polypeptides of the invention are useful as immunoassay
XX reagents in the presymptomatic diagnosis of insulin-dependent type I
XX diabetes mellitus (IDDM). The ability of the polypeptides to bind the

CC antibody binding site on islet cell antibodies (ICAs) also confers
 CC utility in the binding or blocking of human immunoglobulin, T-cells, or B
 CC -cells involved in IDDM. The products can also be used to study the
 CC biological mechanisms involved in islet cell destruction and the
 CC appearance of ICA. The immunoreactivity profile with different antigens
 CC can provide diagnostically significant information concerning the nature
 CC of the disease, e.g. subtypes, the state of the disease, the proximity to
 CC onset of the disease, and the efficacy of therapy, e.g. immune therapy.
 CC The antigens can serve as stimulating antigens for T-cell culture,
 CC permitting significantly improved T-cell cloning, identification, and
 CC growth. The availability of large quantities of pure antigen enables the
 CC development of highly sensitive and specific immunoassays which can be
 CC used to screen the general population of presymptomatic IDDM or a
 CC cell antigen to develop IDDM. This sequence represents the human islet
 CC cell antigen from clone ICA-512
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 24; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 4e-16;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLVKVESSPSSRDYINASPIIEHDP 24
 |||||
 DB 110 KLVKVESSPSSRDYINASPIIEHDP 133

RESULT 2
 AAY49854
 ID AAY49854 standard; protein; 294 AA.
 AC AAY49854;
 XX
 DT 21-JAN-2000 (first entry)
 XX
 DE Human pancreatic islet cell antigen ICA-512 ATCC 40706 protein.
 KW Human; pancreatic islet cell antigen; ICA clone; IDDM; diabetes;
 KW insulin dependent type I diabetes mellitus; diagnosis; detection;
 KW immunoglobulin; T-cell; B-cell; antibody binding.
 XX
 OS Homo sapiens.
 XX
 PN US5981700-A.
 XX
 PD 09-NOV-1999.
 XX
 PF 06-JUN-1995; 95US-00468579.
 XX
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 PR 05-MAY-1994; 94US-00239276.
 XX
 PA (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
 XX
 PI Rabin DU;
 XX
 DR WPI; 1999-633370/54.
 DR N-PSDB; AAZ32341.
 XX
 PT New pancreatic islet cell antigens, useful in the diagnosis of insulin-
 PT dependent (Type I) diabetes mellitus.
 XX
 PS Claim 1; Col 83-86; 61pp; English.
 XX
 CC The present sequence represents a human pancreatic islet cell antigen
 CC (ICA) that binds with antibodies found in the sera of patients afflicted
 CC with insulin-dependent (Type I) diabetes mellitus (IDDM). ICAs from the
 CC present invention are encoded by the DNA insert of a recombinant cloning
 CC vehicle selected from ATCC 40550 (AAZ32333, encoding AAY49847), ATCC
 CC 40553 (AAZ32334, encoding AAY49848), ATCC 40554 (AAZ32335, encoding

CC AAY49849), ATCC 40551 (AAZ32336, encoding AAY49850), ATCC 40552
 CC (AAZ32337, encoding AAY49851), ATCC 40703 (AAZ32338, encoding AAY49852),
 CC ATCC 40704 (AAZ22339, encoding AAY49853), ATCC 40705 (AAZ23340), ATCC
 CC 40706 (AAZ32341, encoding AAY49854) and ATCC 75030 (AAZ32342, encoding
 CC AAY49855). ICA proteins and their peptide fragments can be used in the
 CC diagnosis of IDDM and in detecting or blocking human immunoglobulin, T-
 CC cells or B-cells involved in IDDM
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 24; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 4e-16;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLVKVESSPSSRDYINASPIIEHDP 24
 |||||
 DB 110 KLVKVESSPSSRDYINASPIIEHDP 133

RESULT 3
 AAW80486
 ID AAW80486 standard; protein; 294 AA.
 AC AAW80486;
 XX
 DT 01-FEB-1999 (first entry)
 XX
 DE Islet cell antibody antigen encoded by clone ICA-512.
 XX
 KW Pancreatic islet cell antibody antigen; ICA antigen; diagnosis;
 KW treatment; insulin dependent diabetes mellitus; IDDM; autoantibody;
 KW autoimmune disease; human.
 XX
 OS Homo sapiens.
 XX
 PN US5840836-A.
 XX
 PD 24-NOV-1998.
 XX
 PF 05-MAY-1994; 94US-00239276.
 XX
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 XX
 PA (FARB) BAYER CORP.
 XX
 PI Rabin DU;
 XX
 DR WPI; 1999-034118/03.
 DR N-PSDB; AAV63560.
 XX
 PT Polypeptide antigens from pancreatic islet cells - useful for the
 PT detection and treatment of insulin dependent diabetes mellitus.
 XX
 PS Claim 1; Col 51-54; 47pp; English.
 XX
 CC The present sequence represents a pancreatic islet cell antibody (ICA)
 CC antigen. The proteins are used for the diagnosis and treatment of insulin
 CC dependent (Type I) diabetes mellitus (IDDM). They are recognised by
 CC autoantibodies of the disease as IDDM is now thought to be an autoimmune
 CC disease. They were identified from probing of IDDM serum with human CDNA
 CC libraries expressing protein fragments
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 24; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 4e-16;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLVKVESSPSSRDYINASPIIEHDP 24
 |||||

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 11:13:25 ; Search time 19.0843 Seconds
(without alignments)
83.400 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 KLKVESPSRSDYINASPIIHDP 24

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0
Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.psp:**
 - 2: /cgn2_6/ptodata/1/iaa/5B_COMB.psp:**
 - 3: /cgn2_6/ptodata/1/iaa/6A_COMB.psp:**
 - 4: /cgn2_6/ptodata/1/iaa/6B_COMB.psp:**
 - 5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.psp:**
 - 6: /cgn2_6/ptodata/1/iaa/backfiles.psp:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	246	3	US-08-884-569A-4
2	24	100.0	294	2	US-08-468-576B-18
3	24	100.0	294	2	US-08-468-579B-18
4	24	100.0	294	3	US-08-468-577B-18
5	24	100.0	548	2	US-08-468-576B-19
6	24	100.0	548	2	US-08-468-579B-19
7	24	100.0	548	3	US-08-468-577B-19
8	24	100.0	979	3	US-08-514-213A-2
9	24	100.0	979	4	US-09-015-399-5
10	22	91.7	22	4	US-09-552-802B-41
11	19	79.2	19	4	US-09-552-802B-42
12	17	70.8	17	4	US-09-552-802B-40
13	12	50.0	12	4	US-09-552-802B-39
14	12	50.0	12	4	US-09-552-802B-38
15	11	45.8	11	4	US-09-552-802B-52
16	9	37.5	246	3	US-08-884-569A-3
17	9	37.5	298	3	US-08-811-481-4
18	9	37.5	298	4	US-09-876-527-4
19	9	37.5	376	3	US-08-811-481-7
20	9	37.5	376	4	US-09-876-527-7
21	9	37.5	641	3	US-08-811-481-2
22	9	37.5	723	4	US-09-876-527-2
23	9	37.5	723	3	US-08-548-159-5
24	9	37.5	818	3	US-08-811-481-22
25	9	37.5	818	4	US-09-876-527-22
26	9	37.5	969	2	US-08-548-159-1
27	9	37.5	966	2	US-08-548-159-3

Sequence 2, Appli
Sequence 16, Appl
Sequence 16, Appl
Sequence 6, Appli
Sequence 6, Appli
Sequence 26, Appl
Sequence 26, Appl
Sequence 24, Appl
Sequence 24, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 9, Appli
Sequence 14, Appl
Sequence 9, Appli
Sequence 14, Appl
Sequence 16, Appl
Sequence 16, Appl
Sequence 16, Appl

3 US-08-884-569A-2
3 US-08-811-481-16
4 US-09-876-527-16
1 US-08-015-985-6
4 US-09-280-597-6
2 US-08-685-992-26
2 US-09-144-925-26
2 US-08-685-992-24
2 US-09-144-925-24
2 US-08-685-992-10
2 US-09-144-925-10
2 US-08-685-992-9
2 US-08-685-992-14
2 US-09-144-925-9
2 US-09-144-925-14
2 US-08-685-992-16
2 US-09-144-925-16
1 US-08-201-697-16

ALIGNMENTS

RESULT 1

US-08-884-569A-4
; Sequence 4, Application US/08884569A
; Patent No. 6393326
; GENERAL INFORMATION:
; APPLICANT: CHIANG, MING-KO
; APPLICANT: FLANAGAN, JOHN G.
; TITLE OF INVENTION: RECEPTOR TYROSINE PHOSPHATASE, AND USES RELATED THERETO
; FILE REFERENCE: HMV-020.01
; CURRENT APPLICATION NUMBER: US/06/884,569A
; CURRENT FILING DATE: 1997-06-27
; PRIOR APPLICATION NUMBER: 60/021,040
; PRIOR FILING DATE: 1996-07-02
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Mus sp.
; US-08-884-569A-4

Query Match 100.0%; Score 24; DB 3; Length 246;
Best Local Similarity 100.0%; Pred. No. 6.6e-17;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KLKVESPSRSDYINASPIIHDP 24
Db 19 KLKVESPSRSDYINASPIIHDP 42

RESULT 2

US-08-468-576B-18
; Sequence 18, Application US/08468576B
; Patent No. 5955345
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5

```

SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,576B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.7-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-468-576B-18

```

```

Query Match 100.0%; Score 24; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 7.7e-17;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 KLVKVVSSPSRSDYINASPIIIEHDP 24
DB 110 KLVKVVSSPSRSDYINASPIIIEHDP 133

```

```

RESULT 3
US-08-468-579B-18
Sequence 18, Application US/08468579B
Patent No. 5981700
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,579B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 17-FEB-1989

```

```

FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.5-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-468-579B-18

```

```

Query Match 100.0%; Score 24; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 7.7e-17;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 KLVKVVSSPSRSDYINASPIIIEHDP 24
DB 110 KLVKVVSSPSRSDYINASPIIIEHDP 133

```

```

RESULT 4
US-08-468-577B-18
Sequence 18, Application US/08468577B
Patent No. 6001804
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,577B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 11:39:31 ; Search time 47.4217 Seconds
(without alignments)
164.085 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 KLVKVESSPFRSDYINASPIIEHDP 24

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1370721 seqs, 324215800 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1370721

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	979	14	US-10-038-686-3
2	22	91.7	22	15	US-10-378-479-41
3	19	79.2	19	15	US-10-378-479-42
4	17	70.8	17	15	US-10-378-479-40
5	12	50.0	12	15	US-10-378-479-39
6	12	50.0	12	15	US-10-378-479-38
7	11	45.8	11	15	US-10-378-479-52
8	9	37.5	298	9	US-09-876-527-4
9	37.5	238	14	US-10-124-089-4	
10	9	37.5	376	9	US-09-876-527-7
11	9	37.5	376	14	US-10-124-089-7
12	9	37.5	641	9	US-09-876-527-2
13	9	37.5	641	14	US-10-124-089-2

14	9	37.5	818	9	US-09-876-527-22
15	9	37.5	818	14	US-10-124-089-22
16	9	37.5	1012	9	US-09-876-527-16
17	9	37.5	1012	14	US-10-124-089-16
18	7	29.2	64	14	US-10-029-386-30320
19	7	29.2	65	15	US-10-424-599-219467
20	7	29.2	135	16	US-10-767-701-59696
21	7	29.2	235	15	US-10-087-684-94
22	7	29.2	235	15	US-10-218-779-94
23	7	29.2	235	15	US-10-072-012-819
24	7	29.2	263	15	US-10-087-684-93
25	7	29.2	263	15	US-10-218-779-93
26	7	29.2	264	14	US-10-245-539-6
27	7	29.2	289	14	US-10-314-232-47
28	7	29.2	294	9	US-09-788-626-27
29	7	29.2	301	15	US-10-425-114-54172
30	7	29.2	307	9	US-09-788-626-25
31	7	29.2	309	9	US-09-788-626-11
32	7	29.2	310	9	US-09-788-626-10
33	7	29.2	310	9	US-09-788-626-15
34	7	29.2	312	15	US-10-634-027-6
35	7	29.2	313	9	US-09-788-626-17
36	7	29.2	319	15	US-10-634-027-7
37	7	29.2	322	14	US-10-314-232-11
38	7	29.2	334	14	US-10-314-232-7
39	7	29.2	335	9	US-09-788-626-35
40	7	29.2	336	15	US-10-634-027-4
41	7	29.2	344	16	US-10-408-765A-1670
42	7	29.2	401	14	US-10-314-232-15
43	7	29.2	435	14	US-10-238-075-1224
44	7	29.2	442	9	US-09-925-300-950
45	7	29.2	517	15	US-10-425-114-36999

ALIGNMENTS

RESULT 1
 US-10-038-686-3
 ; Sequence 3, Application US/10038686
 ; Publication No. US20030045467A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Orban, Titamez
 ; TITLE OF INVENTION: AUTORANTIGEN VACCINE
 ; FILE REFERENCE: 10276-067001
 ; CURRENT APPLICATION NUMBER: US/10/038,686
 ; CURRENT FILING DATE: 2002-05-31
 ; PRIOR APPLICATION NUMBER: 60/260,068
 ; PRIOR FILING DATE: 2001-01-05
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 979
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-039-686-3

Query Match 100.0%; Score 24; DB 14; Length 979;
 Best Local Similarity 100.0%; Pred. No. 2.2e-15;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVKVESSPFRSDYINASPIIEHDP 24
 Db 752 KLVKVESSPFRSDYINASPIIEHDP 775

RESULT 2
 US-10-378-479-41
 ; Sequence 41, Application US/10378479
 ; Publication No. US20040006202A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Peakman, Mark
 ; APPLICANT: Chicz, Roman M.

```

; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; TITLE OF INVENTION: CD4+ T LYMPHOCYTES
; FILE REFERENCE: 08191-009002
; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US/09/295,868
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US/60/130,355
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-378-479-41

```

```

Query Match 91.7%; Score 22; DB 15; Length 22;
Best Local Similarity 100.0%; Pred. No. 7.8e-15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 3 KVESSPSSRDYINASPIIEHDP 24
Db 1 KVESSPSSRDYINASPIIEHDP 22

```

```

RESULT 3
US-10-378-479-42
; Sequence 42, Application US/10378479
; Publication No. US20040006202A1
; GENERAL INFORMATION:
; APPLICANT: Peakman, Mark
; APPLICANT: Chiciz, Roman M.
; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; TITLE OF INVENTION: CD4+ T LYMPHOCYTES
; FILE REFERENCE: 08191-009002
; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US/09/295,868
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US/60/130,355
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-378-479-42

```

```

Query Match 79.2%; Score 19; DB 15; Length 19;
Best Local Similarity 100.0%; Pred. No. 6.2e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2 LKVESSPSSRDYINASPII 20
Db 1 LKVESSPSSRDYINASPII 19

```

```

RESULT 4
US-10-378-479-40
; Sequence 40, Application US/10378479
; Publication No. US20040006202A1
; GENERAL INFORMATION:
; APPLICANT: Peakman, Mark
; APPLICANT: Chiciz, Roman M.
; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; TITLE OF INVENTION: CD4+ T LYMPHOCYTES
; FILE REFERENCE: 08191-009002

```

```

; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US/09/295,868
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US/60/130,355
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-378-479-40

```

```

Query Match 70.8%; Score 17; DB 15; Length 17;
Best Local Similarity 100.0%; Pred. No. 5.2e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 KLVESPPSRSDYINAS 17
Db 1 KLVESPPSRSDYINAS 17

```

```

RESULT 5
US-10-378-479-38
; Sequence 38, Application US/10378479
; Publication No. US20040006202A1
; GENERAL INFORMATION:
; APPLICANT: Peakman, Mark
; APPLICANT: Chiciz, Roman M.
; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; TITLE OF INVENTION: CD4+ T LYMPHOCYTES
; FILE REFERENCE: 08191-009002
; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US/09/295,868
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US/60/130,355
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-378-479-38

```

```

Query Match 50.0%; Score 12; DB 15; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 3 KVESSPSSRDYI 14
Db 1 KVESSPSSRDYI 12

```

```

RESULT 6
US-10-378-479-39
; Sequence 39, Application US/10378479
; Publication No. US20040006202A1
; GENERAL INFORMATION:
; APPLICANT: Peakman, Mark
; APPLICANT: Chiciz, Roman M.
; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; TITLE OF INVENTION: CD4+ T LYMPHOCYTES
; FILE REFERENCE: 08191-009002
; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 16:28:53 ; Search time 1574.75 Seconds
(without alignments)
720.721 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 KLVKVESSPSRSDYINASPIIEHDP 24

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9043826

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10783095/runat_26102004_185454_25032/app_query.fasta_1.796

-DB=GenEmbl -QFWT=fastp -SUFFIX=oli.rge -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45

-DOCALIGN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000

-USER=US10783095 @cgn1 1 6148 @runat_26102004_185454_25032 -NCPU=6 -ICPU=3
-NO WMAP -LARGEDUTERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=60
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl:*

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_srs:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Table with 4 columns: Result No., Score, Query Match, Length, DB ID, Description. Contains 4 rows of search results.

Table with 5 columns: Index, Accession, Length, Score, Description. Lists various sequences and their properties.

ALIGNMENTS

Table showing alignment details for AR369579, including sequence, length, and source information.

Table with 4 columns: Alignment Scores, Pred. No., Score, Length. Provides statistical data for the alignment.

US-10-783-095-12 (1-24) x AR369579 (1-1210)

QY 1 LysLeuLysValGluSerSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
Db 527 AAACCTGAAGGTGGAGAGCAGCCCTTCTCGAGCGATTACATCAACGCCAGCCCATTTATT 586
QY 21 GluHisAspPro 24
Db 587 GAGCATGACCCT 598

RESULT 2
AR404158 1210 bp mRNA linear PAT 18-DEC-2003
LOCUS Sequence 30 from patent US 6627735.
ACCESSION AR404158
VERSION AR404158.1 GI:40152191

KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1210)
AUTHORS Kindsvogel,W., Jelinek,L.J., Sheppard,P.O., Hagopian,W.A. and Lagasse,J.M.
TITLE Islet cell antigen 1851
JOURNAL Patent: US 6627735-A 30 30-SEP-2003;
FEATURES Location/Qualifiers
source 1..1210
/organism="unknown"
/mol_type="mRNA"

ORIGIN

Alignment Scores: 1.21e-14 Length: 1210
Pred. No.: 24.00 Matches: 24
Score: 24.00 Mismatches: 0
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-783-095-12 (1-24) x AR404158 (1-1210)

QY 1 LysLeuLysValGluSerSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
Db 527 AAACCTGAAGGTGGAGAGCAGCCCTTCTCGAGCGATTACATCAACGCCAGCCCATTTATT 586
QY 21 GluHisAspPro 24
Db 587 GAGCATGACCCT 598

RESULT 3
AR060650 1413 bp DNA linear PAT 29-SEP-1999
LOCUS Sequence 9 from patent US 5840836.
ACCESSION AR060650
VERSION AR060650.1 GI:5987100

KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1413)
AUTHORS Rabin,D.U.
TITLE Pancreatic islet cell antigens obtained by molecular cloning
JOURNAL Patent: US 5840836-A 9 24-NOV-1998;
FEATURES Location/Qualifiers
source 1..1413
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores: 1.39e-14 Length: 1413
Pred. No.: 24.00 Matches: 24
Score: 24.00 Mismatches: 0

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-783-095-12 (1-24) x AR060650 (1-1413)

QY 1 LysLeuLysValGluSerSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
Db 330 AAACCTGAAGGTGGAGAGCAGCCCTTCTCGAGCGATTACATCAACGCCAGCCCATTTATT 389
QY 21 GluHisAspPro 24
Db 390 GAGCATGACCCT 401

RESULT 4
AR085261 1413 bp DNA linear PAT 01-SEP-2000
LOCUS Sequence 9 from patent US 5981700.
ACCESSION AR085261
VERSION AR085261.1 GI:10012031

KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1413)
AUTHORS Rabin,D.U.
TITLE Pancreatic islet cell antigens obtained by molecular cloning
JOURNAL Patent: US 5981700-A 9 09-NOV-1999;
FEATURES Location/Qualifiers
source 1..1413
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores: 1.39e-14 Length: 1413
Pred. No.: 24.00 Matches: 24
Score: 24.00 Mismatches: 0
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-783-095-12 (1-24) x AR085261 (1-1413)

QY 1 LysLeuLysValGluSerSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
Db 330 AAACCTGAAGGTGGAGAGCAGCCCTTCTCGAGCGATTACATCAACGCCAGCCCATTTATT 389
QY 21 GluHisAspPro 24
Db 390 GAGCATGACCCT 401

RESULT 5
AR094588 1413 bp DNA linear PAT 08-SEP-2000
LOCUS Sequence 9 from patent US 6001804.
ACCESSION AR094588
VERSION AR094588.1 GI:10021642

KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1413)
AUTHORS Rabin,D.U.
TITLE Pancreatic islet cell antigens obtained by molecular cloning
JOURNAL Patent: US 6001804-A 9 14-DEC-1999;
FEATURES Location/Qualifiers
source 1..1413
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 28, 2004, 16:24:37 ; Search time 312.578 Seconds
(Without alignments)
403.055 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 KLVKVESSPSRSDYINASPIIEHDP 24

Scoring table: OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4134886 seqs, 2624710521 residues
Word size: 1
Total number of hits satisfying chosen parameters: 8264151

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlp
-C/cg2_1/USPTO.spool_p/US10783095/runat_26102004_185454_25024/app_query.fasta_1.796
-DB=N_Geneseq_23Sep04 -OPMT=fastcap -SUPFIX=oli.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLCN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095@CGN_1_1_1241@runat_26102004_185454_25024 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_23Sep04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Table with columns: Result No., Score, Match Length, DB ID, Description. Contains 6 rows of search results.

Table with columns: ID, AAT95246 standard; cdna; 1210 BP, AC, AAT95246; DT, 27-MAR-1998 (first entry), DE, Human IA-2/ICAS12 protein tyrosine phosphatase, KW, Islet cell antigen 185i; autoantigen; protein tyrosine phosphatase; insulin-dependent diabetes mellitus; IDDM; diagnosis; therapy; human; IA-2/ICAS12; ds., OS, Homo sapiens., PN, MOS732984-AI., PD, 12-SEP-1997., PF, 05-MAR-1997; 97WO-US003532., PR, 06-MAR-1996; 96US-0012927P., PR, 15-OCT-1996; 96US-0027540P., (ZYMO) ZYMOGENETICS INC. (UNIW) UNIV WASHINGTON., PI, Kindsvogel W, Jelinek L, Sheppard PO, Hagopian W, Lagasse J; PA, AAZ23242 Human pan, AAV63561 Islet cel, AAZ46460 Pancreat, AAS5515 PTPase PT, ADO60053 CRH signa, AB147112 Toxicity, AB139766 Human NS, AAG59970 PTPase PT, AB199237 Mouse isc, AAX87627 Human isl, AAI88754 Human aut, AAI88774 Fusion pr, AAI68773 Fusion pr, AAI4280 Human adu, AAT95226 Human isl, AAT95223 Human isl, AAT67297 Type I di, AAT62228 Human isl, AAT95229 Human isl, AAT95222 Macaque I, AAT88480 Mouse ins, AAT95221 Human isl, AAT95231 Macaque I, AAT67295 Type I di, AAT88478 Human ins, AAT88479 Human ins, AAV4936 Mouse pro, AAT95220 Macaque I, AAT67296 Type I di, AAS87632 DNA encod, AAT95368 Human pro, AAT14680 Human sfc, ADP18660 Human rat, ADP95437 Human BEC, AAS87633 DNA encod, AAS55641 Human sig, ADI55437 Human pol, AAL44150 Barley ye, AAZ88566 BTDV-PAV

ALIGNMENTS

RESULT 1
AAT95246
ID AAT95246 standard; cdna; 1210 BP.
XX AC AAT95246;
XX DT 27-MAR-1998 (first entry)
XX DE Human IA-2/ICAS12 protein tyrosine phosphatase.
XX KW Islet cell antigen 185i; autoantigen; protein tyrosine phosphatase; insulin-dependent diabetes mellitus; IDDM; diagnosis; therapy; human; IA-2/ICAS12; ds.
XX OS Homo sapiens.
XX PN MOS732984-AI.
XX PD 12-SEP-1997.
XX PF 05-MAR-1997; 97WO-US003532.
XX PR 06-MAR-1996; 96US-0012927P.
XX PR 15-OCT-1996; 96US-0027540P.
XX (ZYMO) ZYMOGENETICS INC.
XX (UNIW) UNIV WASHINGTON.
XX PI Kindsvogel W, Jelinek L, Sheppard PO, Hagopian W, Lagasse J;
XX PA AAZ23242 Human pan
XX PA AAV63561 Islet cel
XX PA AAZ46460 Pancreat
XX PA AAZ09923 Human isl

DR WPI; 1997-457535/42.
 XX
 PT Mammalian pancreatic islet cell antigen and related DNA - used to detect
 PT auto:antibodies indicative of insulin-dependent diabetes or pre-
 XX disposition to it.
 XX
 PS Example 3; Page 111; 134pp; English.
 XX
 CC This cDNA sequence comprises the cytoplasmic region of human protein
 CC tyrosine phosphatase IA-2/ICA512. It was obtained from UB7MG glioblastoma
 CC cells (ATCC M85) by PCR amplification (see AAT95244-45). The 1.2 kb
 CC product was ligated into pCEP. Sera from prediabetic subjects were
 CC screened for IA-2/ICA512 autoantibodies. A novel islet cell antigen,
 CC designated 1851 (see AA35296-97), has been identified that forms an
 CC immune complex with an autoantibody found in patients at risk of, or
 CC predisposed to, insulin-dependent diabetes mellitus
 XX
 SQ Sequence 1210 BP; 245 A; 383 C; 370 G; 212 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4,85e-15 Length: 1210
 Score: 24.00 Matches: 24
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-783-095-12 (1-24) x AAT95246 (1-1210)

Qy 1 LysLeuLysValGluSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
 Db 527 AAACCTGAAGTGGAGGAGCCCTTCTCGAGCGATTACATCAACGCCAGCCCATTTATT 586
 Qy 21 GluHisAspPro 24
 Db 587 GAGCATGACCCCT 598

RESULT 2
 AAZ09922
 ID AAZ09922 standard; cDNA; 1413 BP.
 AC AAZ09922;
 DT 29-NOV-1999 (first entry)
 XX
 DE Human islet cell antigen clone ICA-512 cDNA.

KW Islet cell antigen; ICA; human; pancreas; antibody; IDDM; immunoassay;
 KW insulin-dependent type I diabetes mellitus; diagnosis; immunoglobulin;
 KW T-cell; B-cell; islet cell destruction; immunoreactivity; disease;
 KW immune therapy; stimulating antigen; screening; ds.

XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT CDS 1..886
 FT FT /*tag= a
 FT FT /codon_start= 3
 FT FT /note= "partial sequence, no start codon given"
 FT FT /product= "ICA-512"

PN US955345-A.
 XX
 XX PD 21-SEP-1999.
 XX
 XX PF 06-JUN-1995; 95US-00468576.
 XX PR 17-FEB-1989; 89US-00312543.
 XX PR 04-DEC-1989; 89US-00441703.
 XX PR 14-JUN-1991; 91US-00715181.
 XX PR 08-JUN-1992; 92US-00872646.
 XX PR 05-MAY-1994; 94US-00239276.
 XX

PA (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
 XX
 PI Rabin DU;
 XX
 DR WPI; 1999-560454/47.
 DR P-PSDB; AAY33360.
 XX

PT Isolated DNA encoding antigens reactive with serum from diabetics useful
 PT to develop products for the diagnosis and monitoring of insulin-dependent
 XX diabetes mellitus.
 XX
 CC Claim 1; Col 51-56; 63pp; English.

CC This invention describes novel isolated and purified DNAs encoding
 CC pancreatic islet cell antigens reactive with islet cell antibodies found
 CC in sera of patients with insulin-dependent type I diabetes mellitus
 CC (IDDM). The polypeptides of the invention are useful as immunoassay
 CC reagents in the presymptomatic diagnosis of insulin-dependent type I
 CC diabetes mellitus (IDDM). The ability of the polypeptides to bind the
 CC antibody binding site on islet cell antibodies (ICAs) also confers
 CC utility in the binding or blocking of human immunoglobulin, T-cells, or B
 CC -cells involved in IDDM. The products can also be used to study the
 CC biological mechanisms involved in islet cell destruction and the
 CC appearance of ICA. The immunoreactivity profile with different antigens
 CC can provide diagnostically significant information concerning the nature
 CC of the disease, e.g. subtypes, the state of the disease, the proximity to
 CC onset of the disease, and the efficacy of therapy, e.g. immune therapy.
 CC The antigens can serve as stimulating antigens for T-cell culture, and
 CC permitting significantly improved T-cell cloning, identification, and
 CC growth. The availability of large quantities of pure antigen enables the
 CC development of highly sensitive and specific immunoassays which can be
 CC used to screen the general population of presymptomatic IDDM or a
 CC predisposition to develop IDDM. This sequence encodes the human islet
 CC cell antigen from clone ICA-512
 XX

SQ Sequence 1413 BP; 289 A; 459 C; 412 G; 253 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5.61e-15 Length: 1413
 Score: 24.00 Matches: 24
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-783-095-12 (1-24) x AAZ09922 (1-1413)

Qy 1 LysLeuLysValGluSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
 Db 330 AAACCTGAAGTGGAGGAGCCCTTCTCGAGCGATTACATCAACGCCAGCCCATTTATT 389
 Qy 21 GluHisAspPro 24
 Db 390 GAGCATGACCCCT 401

RESULT 3
 AAZ32341
 ID AAZ32341 standard; cDNA; 1413 BP.
 XX
 AC AAZ32341;
 XX
 DT 21-JAN-2000 (first entry)
 XX
 DE Human pancreatic islet cell antigen ICA-512 ATCC 40706 cDNA sequence.

XX Human; pancreatic islet cell antigen; ICA clone; IDDM; diabetes;
 XX insulin dependent type I diabetes mellitus; diagnosis; detection;
 XX immunoglobulin; T-cell; B-cell; antibody binding; ss.
 OS Homo sapiens.
 XX
 XX PN US981700-A.
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 20:02:28 ; Search time 71.1325 Seconds
(without alignments)
239.819 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 KLVKVESSPRSDYINASPIIHDP 24

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues
Word size: 1
Total number of hits satisfying chosen parameters: 1642088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Command line parameters:
-WCDEU=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10783095/runat_26102004.185455.25065/app_query.fasta_1.796
-DB=issued Patents_NA -QMT=rastap -SUFFIX=oli.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=Dits -START=1 -END=1 -MATRIX=oligo -TRANS=numan40.cdd
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-CUTFM=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10783095 @CEN 1 1.213 @runat 26102004.185455.25065 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/FACTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Table with 5 columns: Result No., Score, Query Match, Length, ID, Description. Contains 12 rows of search results.

Table with 5 columns: Index, Score, Query Match, Length, ID, Description. Contains 45 rows of search results.

ALIGNMENTS

RESULT 1
US-08-811-481-30
; Sequence 30, Application US/08811481
; Patent No. 6300093
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Jelinek, Laura J.
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Hagopian, William A.
; APPLICANT: LaGasse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,481
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lindenfelser, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1210 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-811-481-30
Alignment Scores:
Pred. No.: 1.52e-15 Length: 1210
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-10-783-095-12 (1-24) x US-08-811-481-30 (1-1210)
QY 1 LysLeuLysValGluSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
Db 527 AAACGAAAGTGGAGAGCGCCCTTCGGAGCGATTACATCAACGCCGCCCAATTATT 586
QY 21 GluHisAspPro 24
Db 587 GAGCATGACCCT 598

```

```

RESULT 2
US-09-876-527-30
; Sequence 30, Application US/09876527
; Patent No. 6627735
; GENERAL INFORMATION:
; APPLICANT: Jelinek, Wayne
; Jelinek, Laura J.
; Sheppard, Paul O.
; Hagopian, William A.
; LaGasse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/876,527
; FILING DATE: 07-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/911,481
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lingenfelter, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1210 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-09-876-527-30
Alignment Scores:
Pred. No.: 1.52e-15 Length: 1210
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
US-10-783-095-12 (1-24) x US-09-876-527-30 (1-1210)
QY 1 LysLeuLysValGluSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
Db 527 AAACGAAAGTGGAGAGCGCCCTTCGGAGCGATTACATCAACGCCGCCCAATTATT 586
QY 21 GluHisAspPro 24
Db 587 GAGCATGACCCT 598

```

```

RESULT 3
US-08-239-276-9
; Sequence 9, Application US/08239276
; Patent No. 5840836
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/239,276
; FILING DATE: 05-MAY-1994
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.4-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1413 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 28, 2004, 22:07:29 ; Search time 1265.93 Seconds.
(without alignments)
97.211 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 KLVKVESSFRSDYINASPIIEHPD 24
Scoring table: OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3413475 seqs, 2563800928 residues

Word size: 1
Total number of hits satisfying chosen parameters: 6815939

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10783095/runat_26102004_185456_25148/app_query.fasta_1.796
-DB=Published Applications NA -QPMT=fastap -SUFFIX=oli.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNIGS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODELOCAL -OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10783095@cgn_1_1189_@runat_26102004_185456_25148
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

8

Table with columns: Result No., Score, Query Match, Length, DB ID, Description. Contains 45 rows of search results.

ALIGNMENTS

RESULT 1
US-09-876-527-30
; Sequence 30, Application US/09876527
; Patent No. US20020102616A1
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; Jelinek, Laura J.
; Sheppard, Paul O.
; Hagopian, William A.
; LaGasse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS: 34
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; MEDIUM TYPE: Diskette

```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/876,527
FILING DATE: 07-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/811,481
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Lingenfelter, Susan
REGISTRATION NUMBER: P-41,156
REFERENCE/DOCKET NUMBER: 95-36
TELEPHONE: 206-442-6675
TELEFAX: 206-442-6678
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1210 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-09-876-527-30

```

```

Alignment Scores:
Pred. No.: 6.7e-16 Length: 1210
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-10-783-095-12 (1-24) x US-09-876-527-30 (1-1210)

```

```

Qy 1 LvsLeuLysValGluSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
Db 527 AAACGAGGTGGAGGAGCCCTTCGGAGCGATTACATCAAGCCGCCCTTTATT 586
Qy 21 GluHisAspPro 24
Db 587 GAGCATGACCCT 598
RESULT 2
US-10-124-089-30
; Sequence 30, Application US/10124089
; Publication No. US2003016067A1
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; Jelinek, Laura J.
; Sheppard, Paul O.
; Hagopian, William A.
; Lagaesse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESS: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; APPLICATION NUMBER: US/10/124,089
; FILING DATE: 16-Apr-2002

```

```

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/811,481
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Lingenfelter, Susan
REGISTRATION NUMBER: P-41,156
REFERENCE/DOCKET NUMBER: 95-36
TELEPHONE: 206-442-6675
TELEFAX: 206-442-6678
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1210 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-10-124-089-30

```

```

Alignment Scores:
Pred. No.: 6.7e-16 Length: 1210
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0
US-10-783-095-12 (1-24) x US-10-124-089-30 (1-1210)

```

```

Qy 1 LvsLeuLysValGluSerProSerArgSerAspTyrIleAsnAlaSerProIleIle 20
Db 527 AAACGAGGTGGAGGAGCCCTTCGGAGCGATTACATCAAGCCGCCCTTTATT 586
Qy 21 GluHisAspPro 24
Db 587 GAGCATGACCCT 598
RESULT 3
US-10-175-523-167
; Sequence 167, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laerg, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MEPTS)
; FILE REFERENCE: 3235/1J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 167
; LENGTH: 3397
; TYPE: DNA

```

GenCore version 5.1.6
 Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model
 Run on: October 28, 2004, 11:13:10 ; Search time 12.4337 Seconds
 (without alignments)
 185.721 Million cell updates/sec

Title: US-10-783-095-12
 Perfect score: 24
 Sequence: 1 KLKVESPPRSRDYINASPIIEHDP 24

Scoring table: OLIGO
 Gapop 60.0, Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0
 Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
 Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- 1: pir1:*
- 2: pir2:*
- 3: pir3:*
- 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	24	100.0	548	1 I37577	islet cell antigen
2	24	100.0	922	1 S54342	protein-tyrosine-p
3	24	100.0	979	1 JC2349	protein-tyrosine-p
4	24	100.0	996	2 I48721	PTP 35 protein - m
5	9	37.5	248	2 T46903	hypothetical prote
6	9	37.5	1015	2 JC5062	phogrin precursor
7	9	37.5	1015	2 JC5263	transmembrane tyro
8	7	29.2	113	2 F90464	hypothetical prote
9	7	29.2	211	2 G90209	conserved hypothet
10	7	29.2	377	1 A48711	protein-tyrosine-p
11	7	29.2	397	2 T11777	phosphoglycerate t
12	7	29.2	397	2 AI0806	phosphoglycerate t
13	7	29.2	398	2 I56540	protein-tyrosine-p
14	7	29.2	477	2 I38409	adenyl cyclase-a
15	7	29.2	519	2 T48498	hypothetical prote
16	7	29.2	582	2 A57068	protein-tyrosine-p
17	7	29.2	583	2 S17671	protein-tyrosine-p
18	7	29.2	668	2 T34317	protein-tyrosine-p
19	7	29.2	680	2 JC8052	protein tyrosine p
20	7	29.2	699	2 JC6132	protein-tyrosine-p
21	7	29.2	700	1 S20053	protein-tyrosine-p
22	7	29.2	711	1 S28391	protein-tyrosine-p
23	7	29.2	711	2 T45160	protein-tyrosine-p
24	7	29.2	756	2 S67333	hypothetical prote
25	7	29.2	796	1 JC1285	protein-tyrosine-p
26	7	29.2	802	1 A36065	protein-tyrosine-p
27	7	29.2	829	1 A47373	protein-tyrosine-p
28	7	29.2	839	2 C84697	hypothetical prote
29	7	29.2	1118	1 A49724	protein-tyrosine-p

30	7	29.2	1200	2 T43148	probable protein-t
31	7	29.2	1231	2 S53089	protein-tyrosine-p
32	7	29.2	1237	2 A54080	protein-tyrosine-p
33	7	29.2	1290	2 A56493	leucocyte common a
34	7	29.2	1301	1 A41622	protein-tyrosine-p
35	7	29.2	1496	1 A48758	protein-tyrosine-p
36	7	29.2	1499	2 I50212	protein-tyrosine-p
37	7	29.2	1501	2 I58148	protein-tyrosine-p
38	7	29.2	1585	2 T19121	probable protein-t
39	7	29.2	1691	1 D54689	protein-tyrosine-p
40	7	29.2	1863	2 S46217	protein-tyrosine-p
41	7	29.2	1894	2 C54689	protein-tyrosine-p
42	7	29.2	1897	1 TDHULK	leucocyte antigen-
43	7	29.2	1898	2 S46216	leucocyte antigen-
44	7	29.2	1907	2 S50893	protein-tyrosine-p
45	7	29.2	1912	2 A56178	protein-tyrosine-p

ALIGNMENTS

RESULT 1

I37577
 islet cell antigen 512 - human
 N:Alternate names: islet cell autoantigen 3
 N:Contains: protein-tyrosine-phosphatase (EC 3.1.3.48)
 C:Species: Homo sapiens (man)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: I37577; S18121
 R:Rabin, D.O.; Pleasic, S.M.; Shapiro, J.A.; Yoo-Warren, H.; Oles, J.; Hicks, J.M.; Gold J. Immunol. 152, 3183-3188, 1994
 A>Title: Islet cell antigen 512 is a diabetes-specific islet autoantigen related to prot
 A:Reference number: I37577; MUID:94194080; PMID:8144912
 A:Accession: I37577
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-548 <RES>
 A:Cross-references: EMBL:X62899; NID:G32612; PIDN:CAA44688.1; PID:G32613
 C:Genetics:
 A:Gene: GDB:ICA3
 A:Cross-references: GDB:385512
 C:Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphata
 C:Keywords: phosphoprotein; phosphoric monoester hydrolase; transmembrane protein; tyros
 F:196-212/Domain: transmembrane #status predicted <TMN>
 F:346-548/Domain: protein-tyrosine-phosphatase homology #status atypical <PTP>
 F:521/Active site: Cys (phosphocysteine intermediate) #status predicted

Query Match 100.0% Score 24; DB 1; Length 548;
 Best Local Similarity 100.0%; Pred. No. 6.3e-17; Mismatches 0; Indels 0; Gaps 0;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KLKVESPPRSRDYINASPIIEHDP 24
 ||||||||||||||||||||||||||||
 Db 364 KLKVESPPRSRDYINASPIIEHDP 387

RESULT 2

S54342
 N:Alternate names: phosphatase (EC 3.1.3.48), receptor type N precursor - rat
 N:Alternate names: protein-tyrosine-phosphatase IA-2
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: S54342
 R:Kabayashi, Y.; Takahashi, K.; Bardhan, S.; Inagami, T. Biochem. J. 306, 331-335, 1995
 A>Title: Cloning and expression of protein tyrosine phosphatase-like protein derived fro
 A:Reference number: S54342; MUID:95194305; PMID:7887886
 A:Accession: S54342
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-922 <KAM>
 A:Cross-references: EMBL:D38222; NID:G1504061; PIDN:BAA07397.1; PID:G808890
 C:Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphata

C:Keywords: phosphoprotein; phosphoric monoester hydrolase; transmembrane protein; tyrosine phosphatase
F:525-541/Domain: transmembrane #status predicted <TM>
F:677-901/Domain: protein-tyrosine-phosphatase homology <PTP>
F:852/Active site: Cys (phosphocysteine intermediate) #status predicted
F:858/Binding site: substrate phosphate (Arg) #status predicted

Query Match 100.0%; Score 24; DB 1; Length 922;
Best Local Similarity 100.0%; Pred. No. 9.9e-17; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0

QY 1 KLVKVESSPFRSDYINASPIIIEHDP 24
|||||
Db 695 KLVKVESSPFRSDYINASPIIIEHDP 718
|||||

RESULT 3
JC2349
N:Alternate names: protein-tyrosine-phosphatase IA-2
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: JG2349; S40291; I48721
R:Lu, J.; Notkins, A.L.; Lan, M.S.
Biochem. Biophys. Res. Commun. 204, 930-936, 1994
A:Title: Isolation, sequence and expression of a novel mouse brain cDNA, mA-2, and its
A:Reference number: JG2349; MUID:95071416; PMID:7980563
A:Accession: JG2349
A:Molecule type: mRNA
A:Residues: 1-979 <LU>
A:Cross-references: UNIPROT:Q60673; EMBL:U11812; NID:9563737; PIDN:AA52102.1; PID:9563737
A:Experimental source: brain
R:Hendriks, W.; Brugman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.
Submitted to the EMBL Data Library, June 1993
A:Description: Assessment of the expression levels of murine protein-tyrosine phosphatase
A:Reference number: S40280
A:Accession: S40291
A:Molecule type: mRNA
A:Residues: 801-837, 839-852, 'T', 854-907 <HEN>
A:Cross-references: EMBL:Z23060; NID:G438157; PIDN:CAA80595.1; PID:9438158
R:Magistrelli, G.; Covinin, M.; Mosca, M.; Lippoli, G.; Isacchi, A.
Biochem. Biophys. Res. Commun. 217, 154-161, 1995
A:Title: Expression of ptp35, the murine homologue of the PTP2 se-related sequences 1A-2
A:Reference number: I48721; MUID:96095652; PMID:8526904
A:Accession: I48721
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 'MFCARGRSSSPGR', 1-165, 'GDGAGA', 170-362, 'L', 364-614, 'V', 616-674, 'T', 676-858
A:Cross-references: EMBL:X74438; NID:G1089901; PIDN:CAA52453.1; PID:G1089902
C:Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphatase
C:Keywords: brain; glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembrane
F:1-37/Domain: signal sequence #status predicted <SIG>
F:38-979/Product: protein-tyrosine-phosphatase, receptor type N #status predicted <MAT>
F:38-576/Domain: extracellular #status predicted <EXT>
F:696-979/Domain: transmembrane #status predicted <TM>
F:734-958/Domain: intracellular #status predicted <INT>
F:506,524/Binding site: carbonylate (Asn) (covalent) #status predicted
F:909/Active site: Cys (phosphocysteine intermediate) #status predicted
F:915/Binding site: substrate phosphate (Arg) #status predicted

Query Match 100.0%; Score 24; DB 1; Length 979;
Best Local Similarity 100.0%; Pred. No. 1e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1 KLVKVESSPFRSDYINASPIIIEHDP 24
|||||
Db 752 KLVKVESSPFRSDYINASPIIIEHDP 775
|||||

RESULT 4
I48721
P:35 protein - mouse
C:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 23-Jul-1999
C:Accession: I48721
R:Magistrelli, G.; Covinin, M.; Mosca, M.; Lippoli, G.; Isacchi, A.
Biochem. Biophys. Res. Commun. 217, 154-161, 1995
A:Title: Expression of PTP35, the murine homologue of the PTP2 se-related sequences 1A-2
A:Reference number: I48721; MUID:96095652; PMID:8526904
A:Accession: I48721
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-996 <RES>
A:Cross-references: EMBL:X74438; NID:G1089901; PIDN:CAA52453.1; PID:G1089902
C:Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphatase
C:Keywords: phosphoprotein
F:751-975/Domain: protein-tyrosine-phosphatase homology <PTP2>
F:926/Active site: Cys (phosphocysteine intermediate) #status predicted
F:932/Binding site: substrate phosphate (Arg) #status predicted

Query Match 100.0%; Score 24; DB 2; Length 996;
Best Local Similarity 100.0%; Pred. No. 1.1e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1 KLVKVESSPFRSDYINASPIIIEHDP 24
|||||
Db 769 KLVKVESSPFRSDYINASPIIIEHDP 792
|||||

RESULT 5
T46903
N:Protein name: hypothetical protein DKFZP761A0712.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 09-Jul-2004
R:Ansorge, W.; Winkner, U.; Mewes, H.W.; Weil, B.; Wiemann, S.
Submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24134
A:Accession: T46903
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-248 <ANA>
A:Cross-references: UNIPROT:Q9NSR5; EMBL:AL157451
A:Experimental source: adult amygdala; clone DKFZP761A0712
C:Genetics:
A:Note: DKFZP761A0712.1
C:Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphatase

Query Match 37.5%; Score 9; DB 2; Length 248;
Best Local Similarity 100.0%; Pred. No. 0.057;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 11 SDYINASPI 19
|||||
Db 31 SDYINASPI 39
|||||

RESULT 6
JC5062
N:phogrin precursor - human
C:Contains: protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type
C:Species: Homo sapiens (man)
C:Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 09-Jul-2004
C:Accession: JC5062; JC5263; T46903
R:Kawasaki, E.; Hutton, J.C.; Eisenbarth, G.S.
Biochem. Biophys. Res. Commun. 227, 440-447, 1996
A:Title: Molecular cloning and characterization of the human transmembrane protein tyrosin
A:Reference number: JC5062; MUID:97032784; PMID:8878534
A:Contents: islet
A:Accession: JC5062
A:Molecule type: mRNA
A:Residues: 1-1015 <KAW>
A:Cross-references: UNIPROT:Q92932; GB:U66702; NID:G1620663; PIDN:AA50742.1; PID:G162066
R:Smith, P.D.; Barker, K.T.; Wang, J.; Lu, Y.J.; Shipley, J.; Crompton, M.R.
Biochem. Biophys. Res. Commun. 229, 402-411, 1996
A:Title: ICAAR, a novel member of a new family of transmembrane, tyrosine phosphatase-1)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 28, 2004, 19:58:09 ; Search time 2727.04 Seconds
(without alignments)
320.697 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 KLVKVESPRSRDYINASPIIEHPD 24

Scoring table: OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Word size: 1
Total number of hits satisfying chosen parameters: 65641363

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_/UGPTO.spool_p/US10783095/runat_26102004.185454.25046/app_query.fasta_1.796
-DB=EST -QFMT=fastap -SUFFIX=oli.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MCDB=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10783095 @CGN_1_1_9525@runat_26102004.185454.25046 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Table with columns: Result No., Score, Match, Length, DB ID, Description. Contains 9 rows of search results.

Table with columns: 10, 24, 100.0, 547, 5, BQ831314. Contains 45 rows of alignment data.

ALIGNMENTS

BB868512 RIKEN full-length enriched, 16 days neonate male
dienecephalon Mus musculus CDNA clone G630005N01 5', mRNA sequence.

BB868512.1 GI:17114722

EST.
Mus musculus (house mouse)

Mus musculus

DEFINITION
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 330)

Akimura, T., Hirakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,

Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,

Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R.,

Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N.,

Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,

Tanaka, T., Tomaru, A., Toya, T., Watabiki, A., Yasunishi, A.,

Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.

2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-resgsc.riken.jp, URL:http://genome.gsc.riken.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Komno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues

Location/Qualifiers
 1. .330
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strains="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G630005N01"
 /sex="male"
 /tissue_type="diencephalon"
 /dev_stages="16 days neonate"
 /clone_lib="RIKEN full-length enriched, 16 days neonate male diencephalon"

FEATURES
 source

ORIGIN

Alignment Scores: Length: 330
 Pred. No.: 2.88e-14 Matches: 24
 Score: 24.00
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-783-095-12 (1-24) x BB668512 (1-330)

Qy 1 LysLeuYsValGluSerProSerArgSerAspTyIleAsnAlaSerProIlele 20
 Db 222 ARCTGAAAGTGAGAGAGAGCCCTTCGGAGTGATTACATCAACGCCAGCCCATCATC 281
 Qy 21 GluHisAspPro 24
 Db 282 GAGCATGACCCCT 293

RESULT 2
 AW670201
 LOCUS AW670201 357 bp mRNA linear EST 25-APR-2001
 DEFINITION 114090 MARC 1B0V Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AW670201
 VERSION AW670201.1 GI:7526715
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 357)
 AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrnerkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Reitea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keele, J.W.
 TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
 JOURNAL *Genome Res.* 11 (4), 626-630 (2001)

MEDLINE
 PUBMED
 COMMENT

21180013
 11282978
 Contact: Smith TPJ
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACACAGCTATGACCAT
 BACKWARD: GTTTCAGTCCAGTCAGCAGC
 Plate: 111 row: E column: 24
 Seq primer: ATTAGGTGACACTATAG.
 Location/Qualifiers
 1. .357
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH103"
 /clone_lib="MARC 1B0V"
 /note="vector: pCMV SPORT6; Site 1: NotI; Site 2: Sall; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

FEATURES
 source

ORIGIN

Alignment Scores: Length: 357
 Pred. No.: 3.12e-14 Matches: 24
 Score: 24.00
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-783-095-12 (1-24) x AW670201 (1-357)

Qy 1 LysLeuYsValGluSerProSerArgSerAspTyIleAsnAlaSerProIlele 20
 Db 106 AAGCTGAAAGTGAGAGAGAGCCCTTCGGAGTGATTACATCAATGCCAGCCCATATT 165
 Qy 21 GluHisAspPro 24
 Db 166 GAGCATGACCCCT 177

RESULT 3
 CB772195
 LOCUS CB772195 397 bp mRNA linear EST 16-MAY-2003
 DEFINITION AMGNNUC.TRPA3-00011-D3-A trip3 (10298) Rattus norvegicus cDNA clone tip3-00011-d3 5', mRNA sequence.
 ACCESSION CB772195
 VERSION CB772195.1 GI:29860586
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 397)
 AUTHORS Angen EST Program.
 TITLE Angen Rat EST Program
 JOURNAL Unpublished (2003)
 COMMENT Contact: Dan Fitzpatrick
 Angen, Inc
 One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
 Tel: 805 447-4881
 Plate: 00011 row: d column: 3.
 Location/Qualifiers
 1. .397
 /organism="Rattus norvegicus"
 /mol_type="mRNA"

FEATURES
 source

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:57:30 ; Search time 60.7229 Seconds
(without alignments)
227.410 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 24
Sequence: 1 KLVESPSPRSYINASPIIHDHP 24

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 1825181 seqs, 575374646 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1825181
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database : UniProt_02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	358	Q6P3E6	Q6P3E6 mus musculus
2	24	100.0	358	AaH64020	AaH64020 mus musculus
3	24	100.0	591	Q7KZS4	Q7KZS4 homo sapien
4	24	100.0	811	Q9G1A0	Q9G1A0 homo sapien
5	24	100.0	811	AaH07713	AaH07713 homo sapi
6	24	100.0	873	Q925J6	Q925J6 mus musculus
7	24	100.0	876	Q925J7	Q925J7 mus musculus
8	24	100.0	950	Q6NSL1	Q6NSL1 homo sapien
9	24	100.0	950	AaH70053	AaH70053 homo sapi
10	24	100.0	979	PTPN_BOVIN	PTPN_BOVIN
11	24	100.0	979	PTPN_HUMAN	PTPN_HUMAN
12	24	100.0	979	PTPN_MOUSE	PTPN_MOUSE
13	24	100.0	983	PTPN_RAT	PTPN_RAT
14	24	100.0	985	Q6NVE2	Q6NVE2 mus musculus
15	24	100.0	985	AaH68165	AaH68165 mus muscu
16	9	37.5	248	Q9NSR5	Q9NSR5 homo sapien
17	9	37.5	986	Q9V4I9	Q9V4I9 homo sapien
18	9	37.5	998	Q9V4I6	Q9V4I6 homo sapien
19	9	37.5	1001	PTPX_MOUSE	PTPX_MOUSE
20	9	37.5	1004	PTPX_RAT	PTPX_RAT
21	9	37.5	1013	PTPX_MACNE	PTPX_MACNE
22	9	37.5	1015	PTPX_HUMAN	PTPX_HUMAN
23	9	37.5	1042	Q9Y4F8	Q9Y4F8 homo sapien
24	8	33.3	571	Q7S5A7	Q7S5A7 neurospora
25	8	33.3	571	CAR05891	CAR05891 neurospor
26	7	29.2	56	Q6UDH8	Q6UDH8 caris fami
27	7	29.2	62	Q90950	Q90950 gallus gall
28	7	29.2	63	Q90949	Q90949 gallus gall
29	7	29.2	63	Q90951	Q90951 gallus gall
30	7	29.2	113	Q97UX5	Q97UX5 sulfolobus
31	7	29.2	211	Q97ZQ9	Q97ZQ9 sulfolobus

32	7	29.2	229	2	Q845L1	Q845L1 bacillus me
33	7	29.2	240	2	Q9USB5	Q9USB5 schizosacch
34	7	29.2	297	2	Q63476	Q63476 rattus norv
35	7	29.2	352	2	Q8EY48	Q8EY48 debaromyce
36	7	29.2	360	2	Q7SE42	Q7SE42 ashbya goss
37	7	29.2	360	2	AA50599	AA50599 ashbya go
38	7	29.2	374	2	Q8DFJ6	Q8DFJ6 vibrio vuln
39	7	29.2	377	1	PTP2_DICDI	P34138 dictyosteli
40	7	29.2	380	1	CXA7_BRARE	Q92052 brachydanio
41	7	29.2	380	2	Q803S5	Q803S5 brachydanio
42	7	29.2	383	2	Q8MTN0	Q8MTN0 culicoides
43	7	29.2	385	2	Q7MM06	Q7MM06 vibrio vuln
44	7	29.2	397	1	PGTC_SALTY	P37591 salmonella
45	7	29.2	398	2	Q62604	Q62604 rattus norv

ALIGNMENTS

RESULT 1
Q6P3E6
ID Q6P3E6 PRELIMINARY; PRT; 358 AA.
AC Q6P3E6
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE Ptpn protein (Fragment).
GN Name=Ptpn;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Aleksch S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haeh L., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A., Blakesley R.W., Young A.C., Shevchenko Y., Bouffard G.G., Rodriguez A.C., Grimwood J., Green E.D., Dickson M.C., Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE FROM N.A.
STRAIN=FVB/N; TISSUE=Colon;
Submitted (DSC-2003) to the EMBL/GenBank/DBJ databases.
EMBL; BC064020; AaH64020.1; -
InterPro; IPR003595; PTPC motif.
InterPro; IPR000387; TYR_Phosphatase.
InterPro; IPR000242; Tyr_Pp
Pfam; PF00102; Y_phosphatase; 1.
PRINTS; PR00700; PRTYRPHPTASE.
SMART; SM00194; PTEC; 1.
SMART; SMO0404; PTPC motif; 1.
PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
PROSITE; PS50055; TYR_PHOSPHATASE_FTP; 1.
Hydrolase.

DT NON_TER 1 1
 SQ SEQUENCE 358 AA; 40571 MW; 0152828A35992BF2 CRC64;
 Query Match 100.0%; Score 24; DB 2; Length 358;
 Best Local Similarity 100.0%; Pred. No. 2.8e-17;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
 |||||
 131 KLVKVESSPSRSDYINASPIIEHDP 154

Db

RESULT 2
 AAH64020 PRELIMINARY; PRT; 358 AA.
 ID AAH64020
 AC AAH64020
 DT 02-MAR-2004 (TREMBlrel. 27, Created)
 DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
 DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
 DE Ptpn protein (Fragment).
 GN PTPRN
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 SQ SEQUENCE FROM N.A.
 Query Match 100.0%; Score 24; DB 2; Length 591;
 Best Local Similarity 100.0%; Pred. No. 4.5e-17;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
 |||||
 364 KLVKVESSPSRSDYINASPIIEHDP 387

Db

RESULT 4
 Q961A0 PRELIMINARY; PRT; 811 AA.
 ID Q961A0
 AC Q961A0
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE PTPRN protein (Fragment).
 GN Name=PTPRN;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 SQ SEQUENCE FROM N.A.
 Query Match 100.0%; Score 24; DB 2; Length 358;
 Best Local Similarity 100.0%; Pred. No. 2.8e-17;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
 |||||
 131 KLVKVESSPSRSDYINASPIIEHDP 154

Db

RESULT 3
 Q7KZS4 PRELIMINARY; PRT; 591 AA.
 ID Q7KZS4
 AC Q7KZS4
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Protein tyrosine phosphatase, receptor type, N.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 SQ SEQUENCE FROM N.A.
 Query Match 100.0%; Score 24; DB 2; Length 591;
 Best Local Similarity 100.0%; Pred. No. 4.5e-17;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
 |||||
 364 KLVKVESSPSRSDYINASPIIEHDP 387

Db

RESULT 4
 Q961A0 PRELIMINARY; PRT; 811 AA.
 ID Q961A0
 AC Q961A0
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE PTPRN protein (Fragment).
 GN Name=PTPRN;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 SQ SEQUENCE FROM N.A.
 Query Match 100.0%; Score 24; DB 2; Length 358;
 Best Local Similarity 100.0%; Pred. No. 2.8e-17;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVKVESSPSRSDYINASPIIEHDP 24
 |||||
 131 KLVKVESSPSRSDYINASPIIEHDP 154

Db

RESULT 3
 Q7KZS4 PRELIMINARY; PRT; 591 AA.
 ID Q7KZS4
 AC Q7KZS4
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)

GenCore version 5.1.1.6
 Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 06:03:19 ; Search time 84.7229 Seconds
 (without alignments)
 101.620 Million cell updates/sec

Title: US-10-783-095-12
 Perfect score: 123
 Sequence: 1 KLKVVSSPSPSYINASPIIHDHP 24

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

- Database : A_Geneseq_23Sep04:*
 1: Genesecp1980s:*
 2: Genesecp1990s:*
 3: Genesecp2000s:*
 4: Genesecp2001s:*
 5: Genesecp2002s:*
 6: Genesecp2003as:*
 7: Genesecp2003bs:*
 8: Genesecp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	123	100.0	294	2 AAY33360	Aay33360 Human isl
2	123	100.0	294	2 AAY49854	Aay49854 Human pan
3	123	100.0	294	2 AAW80486	Aaw80486 Islet cel
4	123	100.0	294	3 AAY49329	Aay49329 Pancreat
5	123	100.0	379	3 AAY59351	Aay59351 Tyrosine
6	123	100.0	548	2 AAY33361	Aay33361 Human isl
7	123	100.0	548	2 AAY49855	Aay49855 Human pan
8	123	100.0	548	2 AAW80487	Aaw80487 Islet cel
9	123	100.0	548	3 AAY49330	Aay49330 Pancreat
10	123	100.0	548	6 AAW80478	Aaw80478 Human exp
11	123	100.0	548	6 AAW80479	Aaw80479 Human exp
12	123	100.0	548	6 AAW80479	Aaw80479 Human exp
13	123	100.0	548	6 AAW80479	Aaw80479 Human exp
14	123	100.0	591	6 AAW80478	Aaw80478 Human exp
15	123	100.0	591	6 AAW80478	Aaw80478 Human exp
16	123	100.0	591	6 AAW80478	Aaw80478 Human exp
17	123	100.0	916	2 AAR45779	Aar45779 PTPase PT
18	123	100.0	916	6 AAW80478	Aaw80478 Human exp
19	123	100.0	923	7 ADE56165	Ades56165 Rat Prote
20	123	100.0	923	7 ADD45276	Add45276 Rat Prote
21	123	100.0	934	5 ABB06112	Abb06112 Human NS
22	123	100.0	961	2 AAR49039	Aar49039 PTPase PT
23	123	100.0	961	6 AAW80478	Aaw80478 Human exp
24	123	100.0	978	8 AD060054	Ado60054 CRH signa
25	123	100.0	979	2 AAY06606	Aay06606 Human isl

26	123	100.0	979	4 AAG80205	Aag80205 Human aut
27	123	100.0	979	6 ABU04793	Abu04793 Human exp
28	123	100.0	979	6 ABU04784	Abu04784 Human exp
29	123	100.0	979	6 ABU04789	Abu04789 Human exp
30	123	100.0	979	6 ABU04794	Abu04794 Human exp
31	123	100.0	979	7 ADE57230	Ades57230 Human pro
32	123	100.0	979	7 ADE56167	Ades56167 Human pro
33	123	100.0	979	7 ADE57226	Ades57226 Human pro
34	123	100.0	979	7 ADD45278	Add45278 Human pro
35	123	100.0	979	7 ADE57218	Ades57218 Human pro
36	123	100.0	979	7 ADE57222	Ades57222 Human pro
37	123	100.0	996	5 ABB57041	Abb57041 Mouse isc
38	114	92.7	22	3 AAB30279	Aab30279 CD4+ T-ce
39	92	74.8	19	3 AAB30280	Aab30280 CD4+ T-ce
40	91	74.0	298	2 AAW35300	Aaw35300 Human isl
41	91	74.0	370	2 AAW18093	Aaw18093 Type I di
42	91	74.0	376	2 AAW35302	Aaw35302 Human isl
43	91	74.0	641	2 AAW35298	Aaw35298 Macaque i
44	91	74.0	818	2 AAW35297	Aaw35297 Human isl
45	91	74.0	876	2 AAW18091	Aaw18091 Type I di

ALIGNMENTS

RESULT 1
 AAY33360
 ID AAY33360 standard; protein; 294 AA.

AC AAY33360;
 AC AAY33360;
 DT 29-NOV-1999 (first entry)
 DE Human islet cell antigen clone ICA-512 protein fragment.
 KW Islet cell antigen; ICA; human; pancreas; antibody; IDDM; immunoassay;
 insulin-dependent type I diabetes mellitus; diagnosis; immunoglobulin;
 T-cell; B-cell; islet cell destruction; immunoreactivity; disease;
 immune therapy; stimulating antigen; screening.
 OS Homo sapiens.
 PN US9595345-A.
 PD 21-SEP-1999.
 PF 06-JUN-1995; 95US-00468576.
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 PR 05-MAY-1994; 94US-00239276.
 PA (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
 PI Rabin DU;
 PI WPI; 1999-560494/47.
 PI N-PSDB; RAZ09922.
 DR Isolated DNA encoding antigens reactive with serum from diabetics useful
 to develop products for the diagnosis and monitoring of insulin-dependent
 diabetes mellitus.
 XX Example; Col 85-87; 63pp; English.
 CC This invention describes novel isolated and purified DNAs encoding
 pancreatic islet cell antigens reactive with islet cell antibodies found
 in sera of patients with insulin-dependent type I diabetes mellitus
 (IDDM). The polypeptides of the invention are useful as immunoassay
 reagents in the presymptomatic diagnosis of insulin-dependent type I
 diabetes mellitus (IDDM). The ability of the polypeptides to bind the

CC antibody binding site on islet cell antibodies (ICAs) also confers
 CC utility in the binding or blocking of human immunoglobulin, T-cells, or B
 CC -cells involved in IDDM. The products can also be used to study the
 CC biological mechanisms involved in islet cell destruction and the
 CC appearance of ICA. The immunoreactivity profile with different antigens
 CC can provide diagnostically significant information concerning the nature
 CC of the disease, e.g. subtypes, the state of the disease, the proximity to
 CC onset of the disease, and the efficacy of therapy, e.g. immune therapy.
 CC The antigens can serve as stimulating antigens for T-cell culture.
 CC permitting significantly improved T-cell cloning, identification, and
 CC growth. The availability of large quantities of pure antigen enables the
 CC development of highly sensitive and specific immunoassays which can be
 CC used to screen the general population of presymptomatic IDDM or a
 CC predisposition to develop IDDM. This sequence represents the human islet
 CC cell antigen from clone ICA-512
 XX Sequence 294 AA;
 SQ

Query Match 100.0%; Score 123; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 2.1e-11;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVKVESSPSRSDYINASPIIHDHP 24
 |||||
 Db 110 KLVKVESSPSRSDYINASPIIHDHP 133

RESULT 2
 AAY49854
 ID AAY49854 standard; protein; 294 AA.

XX AAY49854;
 XX 21-JAN-2000 (first entry)
 DT Human pancreatic islet cell antigen ICA-512 ATCC 40706 protein.
 DE
 XX Human; pancreatic islet cell antigen; ICA clone; IDDM; diabetes;
 KW insulin dependent type I diabetes mellitus; diagnosis; detection;
 KW immunoglobulin; T-cell; B-cell; antibody binding.
 XX Homo sapiens.
 OS
 XX US5981700-A.
 PN 09-NOV-1999.
 XX 06-JUN-1995; 95US-00468579.
 XX 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 PR 05-MAY-1994; 94US-00239276.
 XX (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
 PA Rabin DU;
 XX WPI; 1999-633370/54.
 DR N-PSDB; AAZ32341.
 XX New pancreatic islet cell antigens, useful in the diagnosis of insulin-
 PT dependent (Type 1) diabetes mellitus.
 XX Claim 1; Col 83-86; 61pp; English.
 CC The present sequence represents a human pancreatic islet cell antigen
 CC (ICA) that binds with antibodies found in the sera of patients afflicted
 CC with insulin-dependent (Type 1) diabetes mellitus (IDDM). ICAs from the
 CC present invention are encoded by the DNA insert of a recombinant cloning
 CC vehicle selected from ATCC 40550 (AAZ32333, encoding AAY49847), ATCC
 CC 40553 (AAZ32334, encoding AAY49848), ATCC 40554 (AAZ32335, encoding

CC AAY49849), ATCC 40551 (AAZ32336, encoding AAY49850), ATCC 40552
 CC (AAZ32337, encoding AAY49851), ATCC 40703 (AAZ32338, encoding AAY49852),
 CC ATCC 40704 (AAZ32339, encoding AAY49853), ATCC 40705 (AAZ32340), ATCC
 CC 40706 (AAZ32341, encoding AAY49854) and ATCC 75030 (AAZ32342, encoding
 CC AAY49855). ICA proteins and their peptide fragments can be used in the
 CC diagnosis of IDDM and in detecting or blocking human immunoglobulin, T-
 CC cells or B-cells involved in IDDM
 XX Sequence 294 AA;
 SQ

Query Match 100.0%; Score 123; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 2.1e-11;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVKVESSPSRSDYINASPIIHDHP 24
 |||||
 Db 110 KLVKVESSPSRSDYINASPIIHDHP 133

RESULT 3
 AAW80486
 ID AAW80486 standard; protein; 294 AA.

XX AAW80486;
 XX 01-FEB-1999 (first entry)
 DT Islet cell antibody antigen encoded by clone ICA-512.
 DE
 XX Pancreatic islet cell antibody antigen; ICA antigen; diagnosis;
 KW treatment; insulin dependent diabetes mellitus; IDDM; autoantibody;
 KW autoimmune disease; human.
 XX Homo sapiens.
 OS
 XX US5840836-A.
 PN 24-NOV-1998.
 XX 05-MAY-1994; 94US-00239276.
 XX 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 XX (FARB) BAYER CORP.
 PA Rabin DU;
 XX WPI; 1999-034118/03.
 DR N-PSDB; AAV63560.
 XX Polypeptide antigens from pancreatic islet cells - useful for the
 PT detection and treatment of insulin dependent diabetes mellitus.
 XX Claim 1; Col 51-54; 47pp; English.
 XX The present sequence represents a pancreatic islet cell antibody (ICA)
 CC antigen. The proteins are used for the diagnosis and treatment of insulin
 CC dependent (Type 1) diabetes mellitus (IDDM). They are recognised by
 CC autoantibodies of the disease as IDDM is now thought to be an autoimmune
 CC disease. They were identified from probing of IDDM serum with human cDNA
 CC libraries expressing protein fragments
 XX Sequence 294 AA;
 SQ

Query Match 100.0%; Score 123; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 2.1e-11;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLVKVESSPSRSDYINASPIIHDHP 24
 |||||

GenCore version 5.1.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:22:24 ; Search time 25.4458 Seconds
(without alignments)
62.550 Million cell updates/sec

Title: US-10-783-095-12
Perfect score: 123
Sequence: 1 KLVKVESSPSRSDYINASPIIHDHP 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Table with 5 columns: Result No., Score, Query Match Length, DB ID, Description. Contains 27 rows of search results.

Table with 5 columns: Line number, Score, Query Match Length, DB ID, Description. Contains 45 rows of search results.

ALIGNMENTS

RESULT 1
US-08-884-569A-4
; Sequence 4, Application US/08884569A
; Patent No. 6399326
; GENERAL INFORMATION:
; APPLICANT: CHIANG, MING-KO
; APPLICANT: FLANAGAN JOHN G.
; TITLE OF INVENTION: RECEPTOR TYROSINE PHOSPHATASE, AND USES RELATED THERETO
; FILE REFERENCE: HMV-020.01
; CURRENT APPLICATION NUMBER: US/08/884.569A
; PRIOR FILING DATE: 1997-06-27
; PRIOR APPLICATION NUMBER: 60/021,040
; PRIOR FILING DATE: 1996-07-02
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Mus sp.
US-08-884-569A-4

Query Match 100.0%; Score 123; DB 3; Length 246;
Best Local Similarity 100.0%; Pred.No. 1.1e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KLVKVESSPSRSDYINASPIIHDHP 24
Db 19 KLVKVESSPSRSDYINASPIIHDHP 42

RESULT 2
US-08-468-576B-18
; Sequence 18, Application US/08468576B
; Patent No. 5955345
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5

```

; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468.576B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.7-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 18:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-468-576B-18

```

```

Query Match 100.0%; Score 123; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 KLVKVESSPSRSDYINASPIIIEHP 24
Db 110 KLVKVESSPSRSDYINASPIIIEHP 133

RESULT 3
US-08-468-579B-18
; Sequence 18, Application US/08468579B
; Patent No. 5981700
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,579B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; APPLICATION NUMBER: US 07/872,646

```

```

; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.5-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 18:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-468-579B-18

```

```

Query Match 100.0%; Score 123; DB 2; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 KLVKVESSPSRSDYINASPIIIEHP 24
Db 110 KLVKVESSPSRSDYINASPIIIEHP 133

RESULT 4
US-08-468-577B-18
; Sequence 18, Application US/08468577B
; Patent No. 6001804
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,577B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989

```


GenCore version 5.1.6
 Copyright (C) 1993 - 2004 CompuGen Ltd.
 OM protein - protein search, using sw model
 Run on: October 28, 2004, 10:54:39 ; Search time 313.735 Seconds
 (without alignments)
 24.802 Million cell updates/sec

Title: US-10-783-095-12
 Perfect score: 123
 Sequence: 1 KLVESFSPRSVDYINASPIIEHDP 24

Scoring table: BLOSUM62
 Gapop 10.0, Gapext 0.5

Searched: 1370721 seqs, 324215800 residues
 Total number of hits satisfying chosen parameters: 1370721

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCR_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

Result No.	Score	Match	Length	ID	Description
1	123	100.0	979	14	US-10-038-686-3
2	114	92.7	22	15	US-10-378-479-41
3	92	74.8	19	15	US-10-378-479-42
4	91	74.0	298	9	US-09-876-527-4
5	91	74.0	288	14	US-10-124-089-4
6	91	74.0	376	9	US-09-876-527-7
7	91	74.0	376	14	US-10-124-089-7
8	91	74.0	641	9	US-09-876-527-2
9	91	74.0	641	14	US-10-124-089-2
10	91	74.0	818	9	US-09-876-527-22
11	91	74.0	818	14	US-10-124-089-22
12	91	74.0	1012	9	US-09-876-527-16
13	91	74.0	1012	14	US-10-124-089-16

SUMMARIES

	82	66.7	17	15	US-10-378-479-40	Sequence 40, Appl
14	59	48.0	12	15	US-10-378-479-38	Sequence 38, Appl
15	59	48.0	12	15	US-10-378-479-39	Sequence 39, Appl
16	55	44.7	11	15	US-10-378-479-52	Sequence 52, Appl
17	49	39.8	213	16	US-10-437-963-115384	Sequence 115384,
18	48	39.0	66	15	US-10-424-599-189618	Sequence 189618,
19	48	39.0	535	14	US-10-032-585-7255	Sequence 7255, Ap
20	48	39.0	609	14	US-10-177-478-9	Sequence 9, Appl
21	46	37.4	112	14	US-10-104-047-2511	Sequence 2511, Ap
22	46	37.4	432	16	US-10-437-963-196840	Sequence 196840,
23	46	37.4	455	9	US-09-793-988-8	Sequence 8, Appl
24	46	37.4	455	15	US-10-429-893-8	Sequence 8, Appl
25	46	37.4	579	14	US-10-369-493-6573	Sequence 6573, Ap
26	46	37.4	587	16	US-10-687-732-25	Sequence 25, Appl
27	45.5	37.0	70	15	US-10-424-599-221903	Sequence 221903,
28	45.5	37.0	390	16	US-10-437-963-131675	Sequence 131675,
29	45	36.6	66	16	US-10-437-963-111307	Sequence 111307,
30	45	36.6	205	16	US-10-437-963-188792	Sequence 188792,
31	45	36.6	312	9	US-09-788-626-24	Sequence 24, Appl
32	45	36.6	312	14	US-10-369-493-15261	Sequence 15261, A
33	45	36.6	312	15	US-10-389-566-2435	Sequence 2435, Ap
34	45	36.6	361	15	US-10-425-114-64408	Sequence 64408, A
35	45	36.6	370	15	US-10-425-114-62602	Sequence 62602, A
36	45	36.6	455	15	US-10-425-114-69763	Sequence 69763, A
37	45	36.6	738	15	US-10-282-122A-77480	Sequence 77480, A
38	45	36.6	1333	15	US-10-621-758A-12	Sequence 12, Appl
39	45	36.6	1333	16	US-10-663-208A-12	Sequence 12, Appl
40	45	36.6	1333	16	US-10-646-301A-12	Sequence 12, Appl
41	45	36.6	1333	16	US-10-736-769-12	Sequence 12, Appl
42	45	36.6	3647	16	US-10-437-963-119793	Sequence 119793,
43	44	35.8	95	9	US-09-925-300-1824	Sequence 1824, Ap
44	44	35.8	201	16	US-10-767-701-32651	Sequence 32651, A
45	44	35.8	201	16	US-10-767-701-32651	Sequence 32651, A

ALIGNMENTS

RESULT 1

US-10-038-686-3 ; Sequence 3, Application US/10038686

Publication No. US200300045467A1

GENERAL INFORMATION:

APPLICANT: Orban, Timar

TITLE OF INVENTION: AUTOANTIGEN VACCINE

FILE REFERENCE: 10276-067001

CURRENT APPLICATION NUMBER: US/10/038,686

PRIOR FILING DATE: 2002-05-31

PRIOR FILING DATE: 2001-01-05

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 979

TYPE: PRT

ORGANISM: Homo sapiens

US-10-038-686-3

Query Match 100.0%; Score 123; DB 14; Length 979;

Best Local Similarity 100.0%; Pred. No. 6.6e-10; Indels 0;

Matches 24; Conservative 0; Mismatches 0;

QV	1	KLVESFSPRSVDYINASPIIEHDP	24
Db	752	KLVESFSPRSVDYINASPIIEHDP	775

RESULT 2

US-10-378-479-41 ; Sequence 41, Application US/10378479

Publication No. US20040006202A1

GENERAL INFORMATION:

APPLICANT: Peakman, Mark

APPLICANT: Chiccz, Roman M.

```

; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; TITLE OF INVENTION: CD4+ T LYMPHOCYTES
; FILE REFERENCE: 08191-009002
; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 09/295,868
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US 60/130,355
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-378-479-41

```

```

Query Match          92.7%; Score 114; DB 15; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 KVESPPSRSDYINASPIIHDHP 24
Db      1 KVESPPSRSDYINASPIIHDHP 22

```

```

RESULT 3
US-10-378-479-42
; Sequence 42, Application US/10378479
; Publication No. US2004006202A1
; GENERAL INFORMATION:
; APPLICANT: Peakman, Mark
; APPLICANT: Chiczo, Roman M.
; TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY DISEASE PROMOTING
; TITLE OF INVENTION: CD4+ T LYMPHOCYTES
; FILE REFERENCE: 08191-009002
; CURRENT APPLICATION NUMBER: US/10/378,479
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US/09/552,802B
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 09/295,868
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US 60/130,355
; PRIOR FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-378-479-42

```

```

Query Match          74.8%; Score 92; DB 15; Length 19;
Best Local Similarity 100.0%; Pred. No. 4.1e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 LKVESPPSRSDYINASPII 20
Db      1 LKVESPPSRSDYINASPII 19

```

```

RESULT 4
US-09-876-527-4
; Sequence 4, Application US/09876527
; Patent No. US20020102616A1
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; Jelinek, Laura J.
; Sheppard, Paul O.
; Hagopian, William A.
; LaGasse, James M.

```

```

; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/876,527
; FILING DATE: 07-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/811,481
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lingenfelter, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 298 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-876-527-4

```

```

Query Match          74.0%; Score 91; DB 9; Length 298;
Best Local Similarity 73.9%; Pred. No. 1.4e-05;
Matches 17; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      2 LKVESPPSRSDYINASPIIHDHP 24
Db      150 LKAEENSHSHSDYINASPIIHDHP 172

```

```

RESULT 5
US-10-124-089-4
; Sequence 4, Application US/10124089
; Publication No. US20030166067A1
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; Jelinek, Laura J.
; Sheppard, Paul O.
; Hagopian, William A.
; LaGasse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2004, 10:55:55 ; Search time 77.5904 Seconds
(without alignments)
129.454 Million cell updates/sec

Title: US-10-783-095-11

Perfect score: 28
Sequence: 1 LAKEWALCAYQAEPTNCAQGGGNK 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database :
- 1: A_Geneseq_23Sep04:*
 - 2: Geneseq1980s:*
 - 3: Geneseq1990s:*
 - 4: Geneseq2000s:*
 - 5: Geneseq2001s:*
 - 6: Geneseq2002s:*
 - 7: Geneseq2003as:*
 - 8: Geneseq2003bs:*
 - 9: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	100.0	294	AAy33360	Human isl
2	28	100.0	294	AAy49854	Human pan
3	28	100.0	294	AAW80486	Islet cel
4	28	100.0	294	AAy49329	Pancreati
5	28	100.0	379	AAy59351	Tyrosine
6	28	100.0	548	AAy33361	Human isl
7	28	100.0	548	AAy49855	Human pan
8	28	100.0	548	AAW80487	Islet cel
9	28	100.0	548	AAy49330	Pancreati
10	28	100.0	548	ABU04788	Human exp
11	28	100.0	548	ABU04790	Human exp
12	28	100.0	548	ABU04792	Human exp
13	28	100.0	548	ABU04791	Human exp
14	28	100.0	591	ABU04783	Human exp
15	28	100.0	591	ABU04785	Human exp
16	28	100.0	591	ABU04795	Human exp
17	28	100.0	923	ADe56165	Rat Prote
18	28	100.0	923	ADd45276	Rat Prote
19	28	100.0	934	ABE806112	Human NS
20	28	100.0	979	AAy06606	Human isl
21	28	100.0	979	AAg80205	Human aut
22	28	100.0	979	AAy04793	Human exp
23	28	100.0	979	ABU04784	Human exp
24	28	100.0	979	ABU04789	Human exp
25	28	100.0	979	ABU04794	Human exp

ID	AAy33360 standard; protein; 294 AA.	28	100.0	979	7	ADe57230 Human Pro
26		28	100.0	979	7	ADe57230 Human Pro
27		28	100.0	979	7	ADe56167 Human Pro
28		28	100.0	979	7	ADe57226 Human Pro
29		28	100.0	979	7	ADd45278 Human Pro
30		28	100.0	979	7	ADe57218 Human Pro
31		28	100.0	979	7	ADe57222 Human Pro
32		24	85.7	24	3	AAb30260 CD4+ T-Ce
33		19	67.9	19	3	AAAB30258 CD4+ T-Ce
34		19	67.9	916	2	AAAR45779 CD4+ T-Ce
35		19	67.9	916	6	AAEU04786 CD4+ T-Ce
36		19	67.9	961	6	AAAR49039 Human exp
37		19	67.9	961	6	AAEU04787 Human exp
38		19	67.9	961	8	AAO60054 CRH signa
39		19	67.9	996	5	ABBS7041 Mouse iLsc
40		18	64.3	18	3	AAAB30259 CD4+ T-Ce
41		17	60.7	17	3	AAAB30257 CD4+ T-Ce
42		16	57.1	16	3	AAy59322 Tyrosine
43		16	57.1	16	3	AAAB30256 CD4+ T-Ce
44		13	46.4	13	3	AAAB30255 CD4+ T-Ce
45		10	35.7	16	3	AAy59337 Tyrosine

ALIGNMENTS

RESULT 1

AAy33360
 XX AAY33360 standard; protein; 294 AA.
 AC AAY33360;
 XX
 DT 29-NOV-1999 (first entry)
 XX
 DE Human islet cell antigen clone ICA-512 protein fragment.
 XX
 KW Islet cell antigen; ICA; human; pancreas; antibody; IDDM; immunoassay;
 KW insulin-dependent type I diabetes mellitus; diagnosis; immunoglobulin;
 KW T-cell; B-cell; islet cell destruction; immunoreactivity; disease;
 KW immune therapy; stimulating antigen; screening.
 XX
 OS Homo sapiens.
 XX
 PN US5955345-A.
 XX
 PD 21-SEP-1999.
 XX
 XX 06-JUN-1995; 95US-00468576.
 XX
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 PR 05-MAY-1994; 94US-00239276.
 XX
 (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
 Rabin DU;
 DR WPI; 1999-560494/47.
 DR N-ESDE; AAZ09922.
 XX
 Isolated DNA encoding antigens reactive with serum from diabetics useful to develop products for the diagnosis and monitoring of insulin-dependent diabetes mellitus.
 PS Example; Col 85-87; 63pp; English.
 XX
 CC This invention describes novel isolated and purified DNAs encoding pancreatic islet cell antigens reactive with islet cell antibodies found in sera of patients with insulin-dependent type I diabetes mellitus (IDDM). The polypeptides of the invention are useful as immunoassay reagents in the presymptomatic diagnosis of insulin-dependent type I diabetes mellitus (IDDM). The ability of the polypeptides to bind the

CC antibody binding site on islet cell antibodies (ICAs) also confers
 CC utility in the binding or blocking of human immunoglobulin, T-cells, or B
 CC -cells involved in IDDM. The products can also be used to study the
 CC biological mechanisms involved in islet cell destruction and the
 CC appearance of ICA. The immunoreactivity profile with different antigens
 CC can provide diagnostically significant information concerning the nature
 CC of the disease, e.g. subtypes, the state of the disease, the proximity to
 CC onset of the disease, and the efficacy of therapy, e.g. immune therapy.
 CC The antigens can serve as stimulating antigens for T-cell culture,
 CC permitting significantly improved T-cell cloning, identification, and
 CC growth. The availability of large quantities of pure antigen enables the
 CC development of highly sensitive and specific immunoassays which can be
 CC used to screen the general population of presymptomatic IDDM or a
 CC predisposition to develop IDDM. This sequence represents the human islet
 CC cell antigen from clone ICA-512
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 28; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 1.2e-21;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAKEWQALCAYQAEPTNCATAGGEGNIK 28
 |||||
 Db 67 LAKEWQALCAYQAEPTNCATAGGEGNIK 94

RESULT 2
 AAY49854
 ID AAY49854 standard; protein; 294 AA.
 XX
 AC AAY49854;
 XX
 DT 21-JAN-2000 (first entry)
 XX
 DE Human pancreatic islet cell antigen ICA-512 ATCC 40706 protein.
 XX
 KW Human; pancreatic islet cell antigen; ICA clone; IDDM; diabetes;
 KW insulin dependent type I diabetes mellitus; diagnosis; detection;
 KW immunoglobulin, T-cell; B-cell; antibody binding.
 XX
 OS Homo sapiens.
 CC
 US9981700-A.
 PN
 PD 09-NOV-1999.
 XX
 PF 06-JUN-1995; 95US-00468579.
 XX
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 PR 05-MAY-1994; 94US-00239276.
 XX
 PA (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.

XX Rabin DU;
 XX
 DR WPI; 1999-633370/54.
 DR N-PSDB; AAZ32341.
 XX
 PT New pancreatic islet cell antigens, useful in the diagnosis of insulin-
 PT dependent (Type I) diabetes mellitus.
 XX
 PS Claim 1; Col 83-86; 61pp; English.
 XX
 CC The present sequence represents a human pancreatic islet cell antigen
 CC (ICA) that binds with antibodies found in the sera of patients afflicted
 CC with insulin-dependent (Type 1) diabetes mellitus (IDDM) ICAs from the
 CC present invention are encoded by the DNA insert of a recombinant cloning
 CC vehicle selected from ATCC 40550 (AAZ32333, encoding AAY49847), ATCC
 CC 40553 (AAZ32334, encoding AAY49848), ATCC 40554 (AAZ32335, encoding

CC AAY49849), ATCC 40551 (AAZ32336, encoding AAY49850), ATCC 40552
 CC (AAZ32337, encoding AAY49851), ATCC 40703 (AAZ32338, encoding AAY49852),
 CC ATCC 40704 (AAZ32339, encoding AAY49853), ATCC 40705 (AAZ32340), ATCC
 CC 40706 (AAZ32341, encoding AAY49854) and ATCC 75030 (AAZ32342, encoding
 CC AAY49855). ICA proteins and their peptide fragments can be used in the
 CC diagnosis of IDDM and in detecting or blocking human immunoglobulin, T-
 CC cells or B-cells involved in IDDM
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 28; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 1.2e-21;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAKEWQALCAYQAEPTNCATAGGEGNIK 28
 |||||
 Db 67 LAKEWQALCAYQAEPTNCATAGGEGNIK 94

RESULT 3
 AAW80486
 ID AAW80486 standard; protein; 294 AA.
 XX
 AC AAW80486;
 XX
 DT 01-FEB-1999 (first entry)
 XX
 DE Islet cell antibody antigen encoded by clone ICA-512.
 XX
 KW Pancreatic islet cell antibody antigen; ICA antigen; diagnosis;
 KW treatment; insulin dependent diabetes mellitus; IDDM; autoantibody;
 KW autoimmune disease; human.
 XX
 OS Homo sapiens.
 XX
 US9840836-A.
 PN
 PD 24-NOV-1998.
 XX
 PF 05-MAY-1994; 94US-00239276.
 XX
 PR 17-FEB-1989; 89US-00312543.
 PR 04-DEC-1989; 89US-00441703.
 PR 14-JUN-1991; 91US-00715181.
 PR 08-JUN-1992; 92US-00872646.
 XX
 PA (FARB) BAYER CORP.
 XX
 PI Rabin DU;
 XX
 DR WPI; 1999-034118/03.
 DR N-PSDB; AAV63560.
 XX
 PT Polypeptide antigens from pancreatic islet cells - useful for the
 PT detection and treatment of insulin dependent diabetes mellitus.
 XX
 PS Claim 1; Col 51-54; 47pp; English.

XX The present sequence represents a pancreatic islet cell antibody (ICA)
 CC antigen. The proteins are used for the diagnosis and treatment of insulin
 CC dependent (Type I) diabetes mellitus (IDDM). They are recognised by
 CC autoantibodies of the disease as IDDM is now thought to be an autoimmune
 CC disease. They were identified from probing of IDDM serum with human cDNA
 CC libraries expressing protein fragments
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 28; DB 2; Length 294;
 Best Local Similarity 100.0%; Pred. No. 1.2e-21;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAKEWQALCAYQAEPTNCATAGGEGNIK 28
 |||||
 |||||